AD-A270 870

April 1993

2

Effects of Practice and Training on the Acquisition and Transfer of Spatial Skills Two Speed-Accuracy Studies

DTIC
S ELECTE D
OCT 18 1993
A

Final Report

Grant AFOSR-91-0367

Submitted by:

David F. Lohman

Lindquist Center for Measurement
The University of Iowa
Iowa City, IA 52242

This document has been approved for public release and sale; its distribution is unlimited

93-24323

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
25 April 1993

3. REPORT TYPE AND DATES COVERED
Final - 30 Sep 91 - 29 Sep 93

4. TITLE AND SUBTITLE
The effects of practice and training on the acquisition and transfer of spatial skills: Two speed-accuracy studies

5. FUNDING NUMBERS
G-AFOSR-91-0367
61102F
2313
A7

6. AUTHOR(S)
David F. Lohman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lindquist Center for Measurement
The University of Iowa
Iowa City, IA 52242

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NL
Building 410
Bolling AFB, DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Two experiments on the effects of practice on the acquisition and transfer of spatial skills were conducted. Both used the methodology of the speed-accuracy study to generate performance curves for each subject in each condition. In the first experiment, subjects practiced either rotating or assembling polygons. They then attempted either a transfer rotation or assembly task that presented either practiced or nonpracticed stimuli. In the second experiment, subjects practiced assembling forms for three sessions in one of four conditions and later were administered the same assembly posttest used in Experiment 1. There were several important findings. First, as expected, practice resulted in substantial improvements in performing the same transformation on the same stimulus set. Second, improvements also transferred to tasks that shared the same procedures. Third, improvements transferred even when procedures varied somewhat between practice and test tasks. Fourth, general improvements in spatial assembly skill were obtained when subjects received practice in assembling concrete materials, but not when they received varied practiced on different assembly tasks. Taken together, these studies suggest caution in interpreting scores on spatial tests for subjects who have had differential experience with the procedures they employ.

14. SUBJECT TERMS
Spatial Ability, Speed-Accuracy, Transfer, Mental Rotation, Mental Assembly

15. NUMBER OF PAGES
51

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	(U)	Unclassified	unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

6 MAY 1993

April 1993

Effects of Practice and Training on the Acquisition and Transfer of Spatial Skills Two Speed-Accuracy Studies

Final Report

Grant AFOSR-91-0367

Submitted by:

David F. Lohman

Lindquist Center for Measurement
The University of Iowa
Iowa City, IA 52242

SUMMARY

Two experiments on the effects of practice on the acquisition and transfer of spatial skills were conducted. Both used the methodology of the speed-accuracy study to generate performance curves for each subject in each condition. In the first experiment, subjects practiced either rotating or assembling polygons. They then attempted either a transfer rotation or assembly task that presented either practiced or nonpracticed stimuli. In the second experiment, subjects practiced assembling forms for three sessions in one of four conditions and later were tested on the same transfer task used in Experiment 1. There were several important findings. First, practice resulted in substantial improvements in performing the same transformation on the same stimulus set. Effect sizes were approximately .8 SD for long exposure trials in both the assembly and rotation tasks. Although substantial, these effects were somewhat smaller than those obtained in studies that did not control for speed-accuracy tradeoff. Second, both experiments showed that improvements also transferred to tasks that shared the same procedures. It was argued that practice clarified task demands, released working memory resources for image construction and retention, and thereby improved the overall efficiency and effectiveness of information processing. Results also showed that procedures need not be identical in practice and test tasks in order to obtain substantial transfer. Subjects who practiced a varied set of paper-and-pencil assembly problems showed an average advantage of .6 SD over control subjects on a computer-based posttest. Furthermore, there was some evidence that general improvements in spatial assembly skill may be better produced through practice with concrete materials than through varied practiced on different assembly tasks. Finally, both experiments showed that subjects who practiced rotating stimuli later assembled or rotated them with greater efficiency than nonpracticed stimuli. However, practice in assembling stimuli produced no such effects. Improvements in processing efficiency (or rate) due to stimulus familiarity were small compared to the much larger improvements in processing effectiveness (or accuracy) due to procedural familiarity. In summary, these studies suggest extreme caution in interpreting scores on spatial tests for subjects who have had differential experience with the procedures they employ. The studies also suggest that practice on spatial tasks is not entirely task-specific; indeed, with a careful match of subject and treatment, general improvements in spatial skills may even be possible.

By
Dist
A-1

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Can spatial abilities be improved with training? On the one hand, effects of practice on spatial tasks are often quite large. For example, Lohman and Nichols (1990) reported gains of .5 to 1.5 SD in number of problems answered correctly on four spatial tests when subjects were retested with the same tests after three days. Krumboltz and Christal (1960) reported average retest gains of .76 SD on one spatial test and 1.05 SD on another spatial test after retest intervals of up to seven hours. Retesting with an alternate form of each test showed the same gain as retesting with the same form.[1] On the other hand, although practice-induced gains on spatial tasks are large, they often do not transfer to tasks that use even slightly different stimuli or require different types of processing (Bethell-Fox & Shepard, 1988; Lohman & Nichols, 1990; Regian & Pellegrino, 1984). For example, Krumboltz and Christal (1960) also retested some subjects on a test known to load on the same spatial factor as the pretest but that used different types of items. These subjects showed no benefit for having taken the pretest. Thus, simple practice effects suggest a more general improvement in spatial skill than has been demonstrated in transfer studies.

The effects of practice on test-like tasks should be an important concern for those who use spatial tests to make decisions about selection and classification. Large practice effects on such tests might invalidate scores for subjects who have fortuitously or deliberately practiced problems similar to those tested. A central issue for both theory and practice, then, is whether practice on one type of spatial task improves performance on other spatial tasks, or if, instead, practice gains are task-specific.

These issues were addressed in two experiments. The purpose of the first experiment was to determine whether practice on a particular type of spatial transformation (rotation or assembly) would produce transferable improvements in performing that same transformation on a different set of stimuli, or instead if practice would result in an improved representation for the practiced stimuli thus producing improved performance on trials requiring that a different transformation be performed on that same stimulus set. The purpose of the second experiment was to investigate the specificity of the procedural transfer observed in the first experiment. In particular, we sought to estimate the extent to which transfer depended on equivalence of task formats and procedures. We also investigated whether different types of practice might aid subjects in acquiring more general spatial skills. Some theories of skill acquisition (e.g., Anderson, 1983) claim that general skills are developed through practice of specific skills in varied contexts. Other theories emphasize the internalization of an external activity in the development of general spatial skills (Vygotsky,1962). These predictions were contrasted in the second experiment.

Effects of Practice on Spatial Tests and Tasks

Studies of the effects of practice and training on the development of spatial abilities vary in the explicitness of the training provided. Least explicit are studies that examine the experiential correlates of spatial abilities. For example, Paterson, Elliot, Anderson, Toops, and Heidbreder (1930) found that boys who reported more experience with mechanics, carpentry, and related activities had higher spatial abilities than similar boys who reported less experience. However, it is unclear whether differential experience was the cause or the consequence of different levels of spatial abilities. Slightly more explicit are studies that examine the effects of a particular course of instruction such as geometry (Brown, 1954; Ranucci, 1952), engineering drawing (Churchill, Curtis, Coombes, & Hassell, 1942; Faubian, Cleveland, & Hassell, 1942), or

[1] Indeed, the susceptibility of spatial and other performance tasks to practice effects was the main reason why Thorndike, Bregman, Cobb, and Woodyard (1926) rejected Stern's (1914) suggestion to measure intelligence using novel tasks. Instead, Thorndike and his collaborators inferred intellectual competence from performance on relatively familiar school tasks.

orthographic drawing (Lajoie, 1986). Effects of such instruction are often large for the specific skills trained, but usually small or nonsignificant on transfer tests of spatial abilities. However, Balke-Aurell's (1982) study of the effects of tracking in the Swedish secondary school system suggests that the cumulative effect of differential educational experiences can be quite large. She found that changes in the relative development of verbal and spatial abilities were strongly related to the type of educational and occupational experiences of students. Those educated in a verbally-oriented curricula showed greater growth in verbal abilities, whereas those educated in a technical curricula showed greater growth in spatial and technical abilities.

More controlled intervention studies have examined the effects of a particular set of training exercises on related spatial tasks. For example, subjects have been trained in rotating concrete objects (Rovet in Olsen & Bialystok, 1983) or in assembling puzzles (Embretson, 1987), have watched films that demonstrated paper-folding (Kyllonen, Lohman, & Snow, 1984) or surface-development (Salomon, 1974) tasks, or have been taught strategies for remembering and classifying spatial transformations (Kyllonen et al., 1984). Performance on related spatial tests or tasks is then examined.

Results of these intervention studies are not easily summarized. A study by Kyllonen et al. (1884) shows why. These investigators compared the effects of treatments designed to teach either an analog strategy on an analytic strategy for solving paper-folding items with a control condition in which subjects were simply given practice with feedback. Items varied systematically in ways known to influence their difficulty or in ways that were expected to influence how they were solved. For example, parts of some items were presented successively (first one fold, then another fold, then the hole punch, then the response alternatives) whereas other items showed all parts of the item simultaneously. On some items subjects selected their answer from a set of options whereas on others they were required to draw their answers. In general, high spatial subjects did best when given practice with feedback, whereas low-spatial subjects did best when given more explicit instructions. However, the effects of treatments depended both on the trial type and the ability profile of the subject.

This last point is particularly troublesome. In fact, training programs often impair the performance some subjects. For example, attempts to teach subjects how to solve three-dimensional rotation problems (Rovet, in Olson & Bialystok, 1983), paper-folding problems (Kyllonen et al., 1984), surface development problems (Salomon, 1974), orthographic projection problems (Lajoie, 1987), and mental assembly problems (Ackerman & Lohman, 1990, Experiment 1) have been found to be disruptive for high-verbal subjects. This is counterintuitive since one would not expect treatments for improving spatial problem solving to be moderated by verbal abilities. One possibility is that these treatments impose or induce external regulation of performance. High ability subjects generally perform better in treatments that allow self-regulation than in treatments that impose external regulation (van der Sander & Schouten, 1985). Low ability subjects show the opposite pattern. External regulation may serve to compensate for the inadequate self-regulation activities low verbal subjects, but interfere with the self-regulation activities of high-verbal subjects.

In summary, intervention studies show that although spatial abilities may be improved with training, results depend on many factors. One reason for confusion may be that such studies do not make explicit which aspects of performance are improved with training, and, of those that are improved, which show transfer to nontrained tasks.

These concerns have been addressed at least in part in the most restricted type of study. In these studies, subjects have been given practice with (or without) feedback on one or more spatial tasks (Alderton, Pellegrino, & Lydiatt, 1984; Levine, Schulman, Brahlek, & Fleishman, 1980), often with the aim of discovering which aspects of performance improve with practice (Lohman & Nichols, 1990; Regian & Pellegrino, 1984). Although transfer is sometimes examined, the primary goal is to investigate practice-induced changes in the task itself. Changes in the predictive validities of scores have also been investigated in this paradigm (Ackerman & Lohman, 1990; Mittelholtz, 1985). Results of these studies have been more consistent (although not always any easier to interpret) than intervention studies. Some of the main findings are:

1. Substantial improvements in speed of response.

Practice on spatial tasks invariably produces faster responding. Latencies usually decline exponentially, thus showing the greatest change in early trials. Effects are largest for relatively simple tasks in which trials are similar. For example, Lohman and Nichols (1990) reported gains of 1.2 - 1.6 SD in number correct on a speeded rotation test when subjects were retested on the same test after a 3-day delay. Gains were almost entirely due to an increase in the number of problems attempted, not to a decrease in errors. Similarly, Ackerman and Lohman (1990) reported average gains of .98 and 1.11 SD in study time and comparison time on four spatial tasks across three practice sessions. These changes are not short-lived. Practice-induced gains persisted unchanged for at least 30 days in one study (Ackerman & Lohman , 1990, Experiment 2) and for two to three months in another study (Alderton et al., 1984).

Although durable, changes are not unambiguous. There is evidence that at least some of the improvement in response speed may be due to changes in speed-accuracy tradeoff. For example, Lohman and Nichols (1990) found that subjects not only solved significantly more rotation items on the posttest than on the pretest, but also made more errors of commission on the posttest. Likewise, Ackerman and Lohman (1990, Experiment 2) found that a large decrease in latencies across sessions was associated with a smaller but significant increase in errors.[2]

The locus of the improvement is also problematic. Regian and Pellegrino (1984) gave groups of low and high spatial ability subjects five sessions of practice in rotating two sets of polygons. One set (Set X) was practiced in all five sessions, the second (Set Z) in Sessions 3 through 5 only. In the analyses, latency was regressed on angular separation between stimuli, separately for each stimulus set, session, and ability group. According to Shepard and Metzler (1971), the slope of the regression estimates rate of rotation whereas the intercept estimates time for all other processes. As expected, both intercepts and slopes for stimulus Set X declined exponentially over the five sessions. (See Ackerman & Lohman, 1990, Experiment 1, for a similar result on two rotation and two assembly tasks.) The important question, however, was whether improvements over the first three sessions with stimulus Set X would transfer to stimulus Set Z when these stimuli were introduced in Session 3. Results showed that those changes reflected in the intercept of the function did transfer to the new stimulus set whereas those reflected in the slope did not transfer. Bethell-Fox and Shepard (1988) report a similar finding. Such results suggest that practice-induced changes in rate of rotation are stimulus specific. However, steeper slopes do not necessarily imply faster rotation. Subjects may make several attempts to rotate unfamiliar stimuli (especially on trials requiring more rotation), or may perform the rotation iteratively on parts of the stimulus. Familiar stimuli may be rotated holistically (Bethell-Fox & Shepard, 1988). Thus, improvements in rate of rotation with practice may reflect changes in how stimuli are represented rather than changes in the rotation operation itself. Therefore, the question of what transfers after training is still somewhat vexed. In particular, it is not clear how much improvement is due to increases in stimulus familiarity, how much to increases in the speed or accuracy of executing particular

[2] Transfer effects reported in some training studies (e.g., Kyllonen et al., 1984) may also reflect changes in speed-accuracy tradeoff induced by contextual factors such the pace of the training procedure.

operations, or how much to improved efficiency in executing a series of repeatedly executed operations.

2. Removal of (possibly irrelevant) sources of difficulty.

Lohman and Nichols (1990), Kyllonen, Lohman, and Snow (1984), and Embretson (1987) all found that practice with feedback eliminated or attenuated the effects of particular sources of difficulty. For example, Kyllonen et al. (1984) concluded that, other than a general improvement in performance, the main effect of practice with feedback on a paper folding task seemed to be to enable subjects to distinguish between a 1/3 fold of the paper and a 1/4 fold. In other words, without feedback, subjects may have ignored this slight difference in line placement, or may have attributed it to poor artwork. With feedback, they learned to attend to the difference. Similarly, Lohman and Nichols (1990) found that the main effect of practice with feedback on a synthesis task was that subjects learned to attend more carefully to the orientation of the test probe. Subjects in the feedback condition of their experiment made fewer errors than those in the no-feedback condition only on negative trials on which the foil was a reflected version of the correctly synthesized figure. Similarly, Embretson (1987) found that practice with feedback on cardboard cutouts eliminated one source of item difficulty on a computer-administered form board test for the group receiving practice. Thus, practice with feedback may increase construct-relevant variance in test scores by reducing variance due to irrelevant sources of difficulty.

3. Systematic changes in correlations between task dependent variables and other variables.

Abilities measured by tasks often change as subjects practice those tasks. For example, Mittelholtz (1985) administered a 144-trial spatial synthesis task without feedback three times to 70 subjects. Correlations between a reference spatial ability factor and errors on the synthesis task declined across sessions (**r** = -.41, -.30, and -.26, for Sessions 1, 2, and 3, respectively), whereas correlations for synthesis latencies increased slightly (**r** = -.28, -.35, and -.37), and correlations for comparison latencies increased the most (**r** = -.44, -.60, -.57). Feedback on the correctness of each response accelerates changes due to practice (Lohman & Nichols, 1990), and may actually improve the predictive validities of task scores. Indeed, Ackerman and Lohman (1990, Experiment 3) found that feedback dramatically improved the multiple correlations between latency and error scores from four spatial tests and number of planes successfully landed in a simulated Air Traffic Control (ATC) task. Spatial tests best predicted performance early in learning on the ATC task for subjects who had one or two sessions of practice on the spatial tests, but best predicted performance late in practice on the ATC task for subjects who had three sessions of practice on the spatial tests. Thus, predictive validities of spatial tests depend on amount of practice given on both the spatial tests and the criterion task, the nature of the criterion task, the presence or absence of feedback, and the nature of the dependent variables used to summarize performance. Of these, the nature of the dependent measure is perhaps the factor most commonly overlooked.

Accuracy-Latency Performance Functions

Previous investigations of the effect of practice on spatial tasks have focused on one aspect of performance -- either response latency or response correctness -- while ignoring or attempting to hold constant the other aspect of performance. This creates several problems. First, for each subject, response accuracy is generally related to response latency through a negatively accelerated exponential function. This means that, when subjects respond cautiously, small changes in response accuracy are associated with large changes in response latency. When subjects respond hastily, however, small changes in response latency are associated with large changes in response accuracy. Thus, changes in performance (indexed by accuracy, latency, or some combination thereof) for individual subjects are difficult to compare unless all adopt the same tradeoff between speed and accuracy, and all maintain this tradeoff as they practice the task.

Practice affects the speed-accuracy tradeoff subjects adopt. Subjects often become more hasty with practice, but differentially so. Investigators have sometimes attempted to adjust scores for individual differences in speed-accuracy tradeoff, usually by subtracting an estimate of the number of problems answered correctly by guessing from the total number of problems answered correctly. Unfortunately, adjusted scores may still not be comparable for different subjects because the within-subject relationship between response accuracy and response latency is nonlinear (see Figure 1) even though the between-subjects relationship may be approximately linear. More importantly, the curve itself differs systematically across subjects (in intercept, curvature, and asymptote)[3]. This last point is particularly troublesome. It means that attempts to hold one aspect of performance constant while examining the other are probably misguided. If speed-accuracy curves for different subjects have different intercepts, different curvatures, and different asymptotes, then what point would represent the same speed-accuracy tradeoff on all curves? One possibility would be to use some constant fraction (say 90%) of asymptotic performance. This requires that one first estimate asymptotic performance, and then move back on the curve until accuracy has stabilized at 90% of its asymptotic value. But why estimate only a portion of the accuracy-latency tradeoff curve rather than the full curve for each subject? Indeed, several investigators have pointed out that estimating how response correctness varies as a function of response latency for each subject and condition can provide a much more comprehensive picture of performance than can a conventional analysis that attends only to one coordinate of one point on this curve or response surface (Wickelgren, 1977; Dosher, 1984), especially when individual differences are the object of inquiry rather than something relegated to the error term (Lohman, 1989). Thus, *we move from the problem of controlling for speed-accuracy tradeoff to the possibility of generating response surfaces that provide a more comprehensive picture of performance.*

Cascade Model

The idea behind the speed-accuracy tradeoff (SAT) function is relatively straightforward. One varies stimulus exposure over a broad range and plots the probability of a correct response (or d' or some other measure of response strength) against total time (exposure latency plus response latency) at each of several levels of stimulus exposure. Figure 1 shows the sort of curve that typically would well fit such data. An exponential function of the form:

$$d' = \lambda\left(1 - e^{-\beta(t-\partial)}\right)$$

is then fitted to the data. d' is the estimated response strength (or accuracy), λ is the asymptote of the curve (or the level of accuracy achieved under liberal time allotment), β its curvature (or the rate of improvement in response accuracy for each additional unit of time), and ∂ its intercept (or the point at which accuracy rises above the level of random responding).

[3] Individual differences in the asymptotes of the curve are particularly common, often quite large, and thus can enormously complicate the task of estimating from a single pair of observations the speed-accuracy tradeoff a subject has adopted. Most tasks of even moderate complexity fit this description, as do some fairly simple ones [such as when college students must compare two three-letter strings (see Lohman, 1989)].

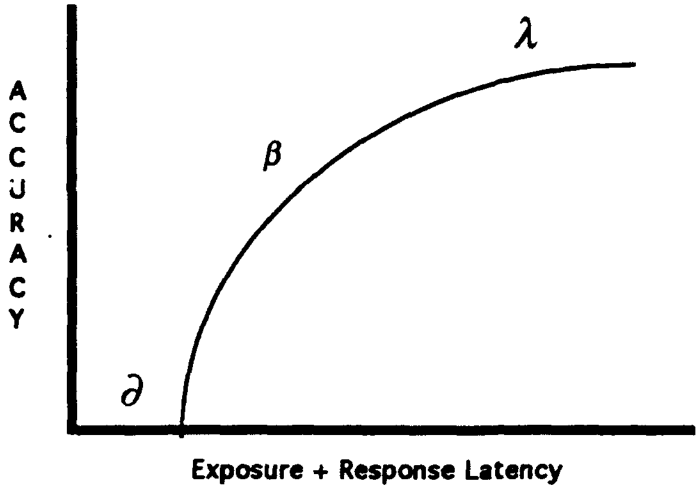


Figure 1. Idealized accuracy-latency (or speed-accuracy) curve.

Interpretation of accuracy-latency performance curves is a bit more complicated than interpretation of univariate scores. In general, such curves better conform to theories of cognition that posit a gradual accumulation of information about the response to be made than do theories that posit discrete outputs of different component operations. For example, McClelland's (1979) Cascade model makes specific predictions about how variations in different aspects of cognition will be reflected in the form of the accuracy-latency performance curve[4]. These predictions are shown in Figure 2.

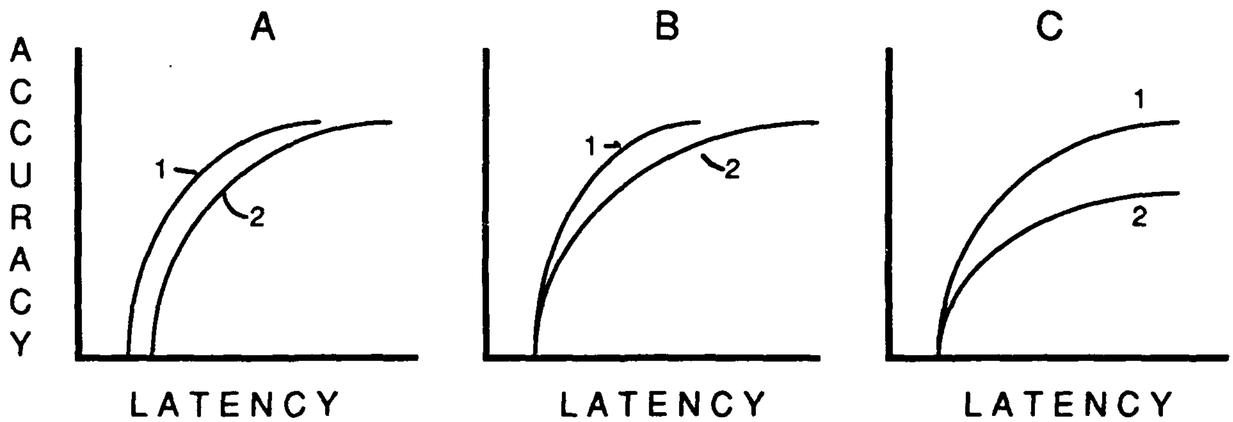


Figure 2. Predictions of the Cascade model. Panel A shows the removal of a fast process in Condition 1, Panel B shows a reduction in rate of execution of a slow process in Condition 1, and Panel C shows an increase in stimulus familiarity in Condition 1.

The location of the intercept of the curve (or the point at which it begins to rise above the level of random responding) reflects the time taken by relatively fast processes in the system. Adding a new fast process simply shifts the curve to the right. Developing automaticity for one or more fast processes shifts the curve to the left. The curvature, on the other hand, is dominated

[4] The Cascade model is probably better viewed as a metaphor for a class of models than as a statement of a particular model. The performance function predicted by the model can assume a variety of shapes since the number of parameters is large. Here I follow McClelland's (1979) exposition, although other predictions are possible.

by the slowest process in the system. Increasing the rate at which this process is executed makes the curve rise more steeply; decreasing its rate flattens the curve. Finally, the asymptote of the curve reflects final level of activation achieved. This depends on factors such as initial activation (primed cognitive units will show higher asymptotes), stimulus quality (degraded stimuli will show lower asymptotes), and stimulus identity (well-learned units will attain higher asymptotes). Because of its specificity, McClelland's (1979) model will be used to generate predictions about the effects of practice in these studies. First, however, it is necessary to review a previous experiment in this series.

The Ackerman-Lohman Study

The first experiment was a revision of a previous study (Ackerman & Lohman, 1990, Experiment 1) that was conducted as part of a series of studies investigating the effects of practice on the validity of spatial tests . The purpose of the Ackerman-Lohman study was to determine whether practice with feedback on different spatial tasks would produce (a) task-specific improvements in performance, or (b) improvements in the ability to perform a particular mental transformation (such as rotation or synthesis), or (c) improvements in the ability to perform other transformations on the practiced stimuli. The order of events in that experiment is shown in Figure 3.

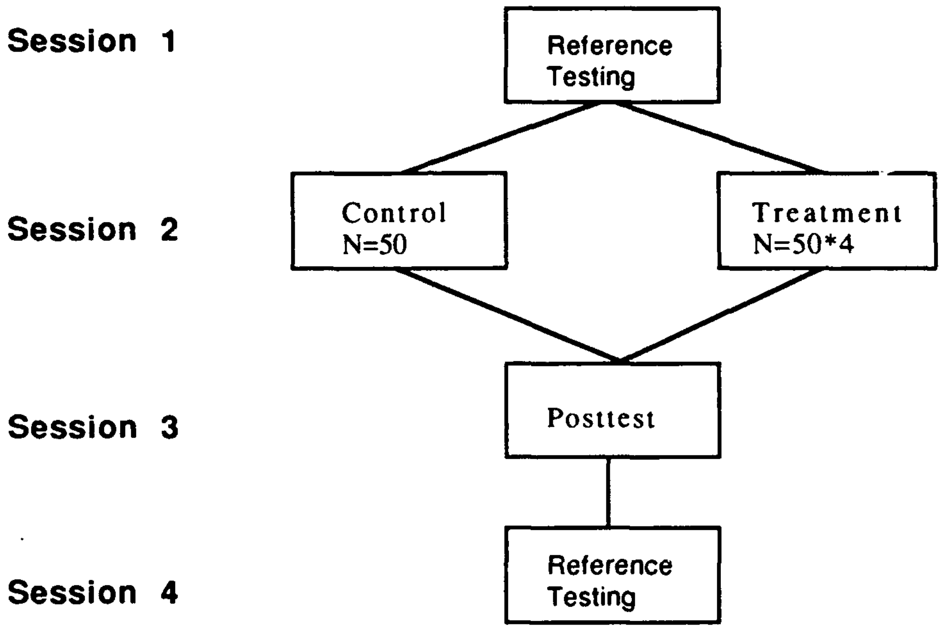


Figure 3. Order of events in the Ackerman-Lohman (1990) study.

<u>Procedure</u>. Subjects were first given a battery of reference tests. They were then randomly assigned to one of four practice treatments or to a no-practice control. Subjects in the treatment groups practiced either rotating or assembling either alphanumeric or polygon stimuli. Each subject attempted 180 trials with dynamic feedback after each incorrect response showing how objects could be combined or rotated. The four practice conditions are shown in Figure 4. Subjects then attempted all four tasks in a posttest. Order of tasks within the posttest was counterbalanced across subjects. Each of the four posttest tasks contained 10 practice trials, 30 self-paced trials, and 150 forced-paced trials. On forced-paced trials, the problem stem was presented for one of six exposures. The goal was to adjust the shortest exposure such that response accuracy would be only slightly better than random responding and to adjust the longest exposure such that accuracy would be near asymptote. Exposures were estimated anew for each subject based on the subject's median response time on similar trials in the self-paced condition.

A test stimulus was then presented on the right side of the screen immediately following the disappearance of the problem stem from the left side of the screen. Subjects were required to accept or reject the test stimulus within a fixed interval. This interval was based on the subject's response times in the self-paced condition.

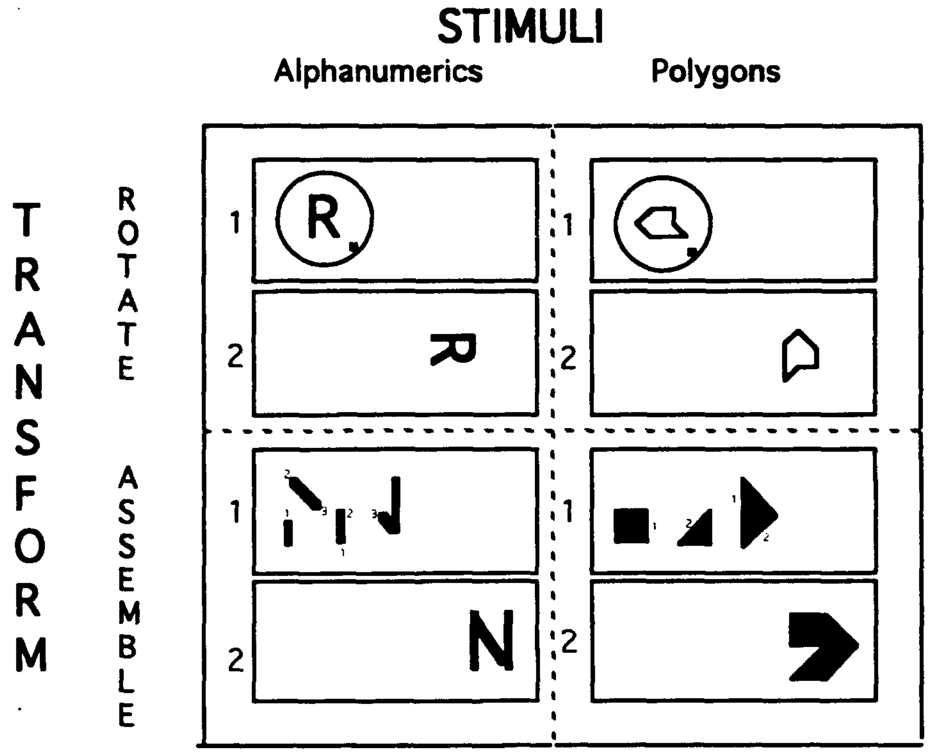


Figure 4. Between-subjects design for the four practice conditions in Session 2 of the Ackerman-Lohman (1990) study. Also the within-subjects design for the posttest administered to all subjects in Session 3. Note that the study stimuli and test stimuli were presented successively.

Results. Overall results for self-paced trials are shown in Figure 5. As expected, subjects showed most improvement when the posttest task was the same as the practiced task. Procedural transfer was next. Subjects who had practiced rotating (or assembling) one stimulus set showed positive transfer when required to perform the same transformation on a different stimulus set. Unexpectedly, stimulus transfer appeared to be negative. Subjects who practiced one transformation on a stimulus set showed negative transfer when required to perform a different transformation on the same stimulus set. Perhaps stimuli were linked to the operations performed on them, which subsequently interfered with attempts to perform different transformations of those stimuli. Finally, subjects who received no practice performed worst of all. Differences between these subjects and subjects in the different-task, different-stimulus group may have reflected unfamiliarity with general task procedures. Note that these differences were primarily in response latency, as would be expected by such a hypothesis

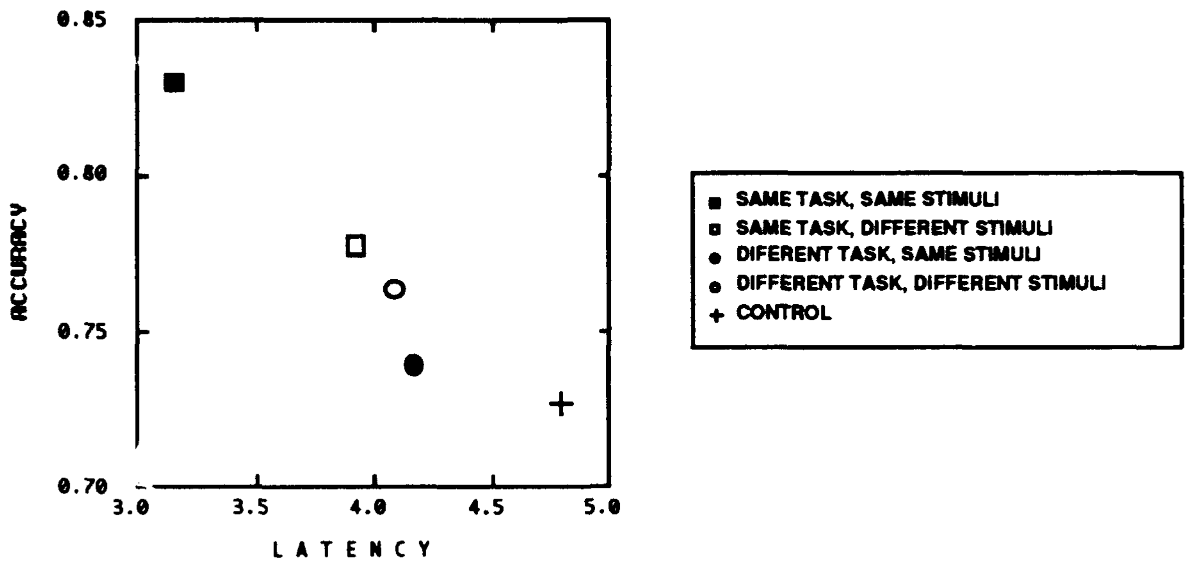


Figure 5. Mean accuracy and mean total latency (study time plus response time) for each of the five practice conditions in the Ackerman-Lohman (1990) study.

However, this orderly pattern in which differences between conditions were in the same direction (and often of approximately the same magnitude) for latency and for accuracy was not observed within conditions. For example, average data for the two assembly posttests is shown in Figure 6. Here subjects who practiced the same (i.e. assembly) task during the practice phase showed higher accuracies but longer latencies when assembling new stimuli than did subjects who practiced the rotation task. Even more complex reversals are evident in the results for particular tasks. These are shown in Figure 7.

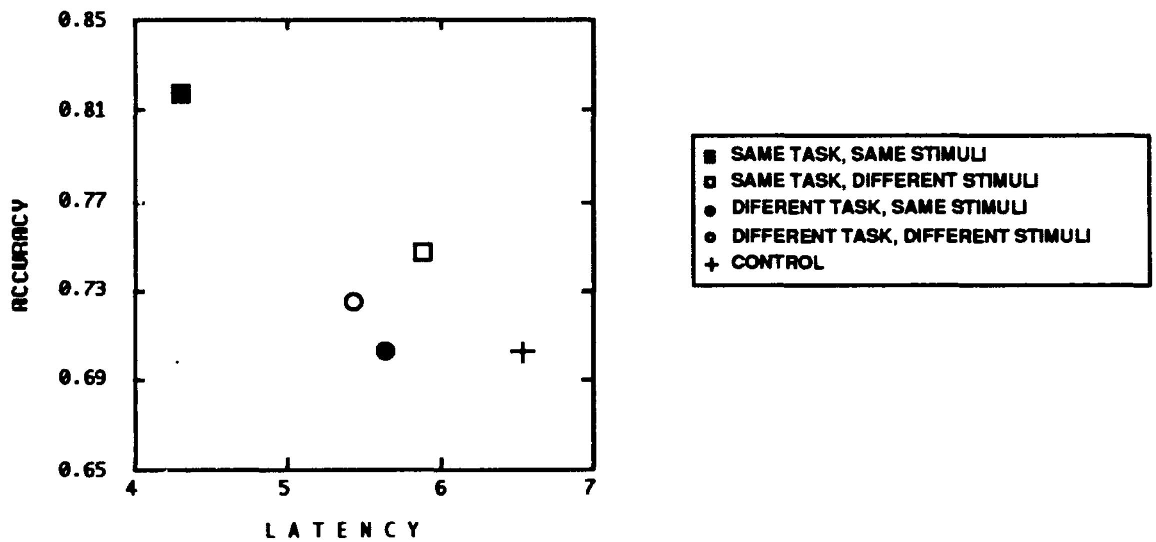


Figure 6. Mean accuracy and mean total latency (study time plus response time) for the two assembly posttests, by practice condition (after Ackerman & Lohman, 1990).

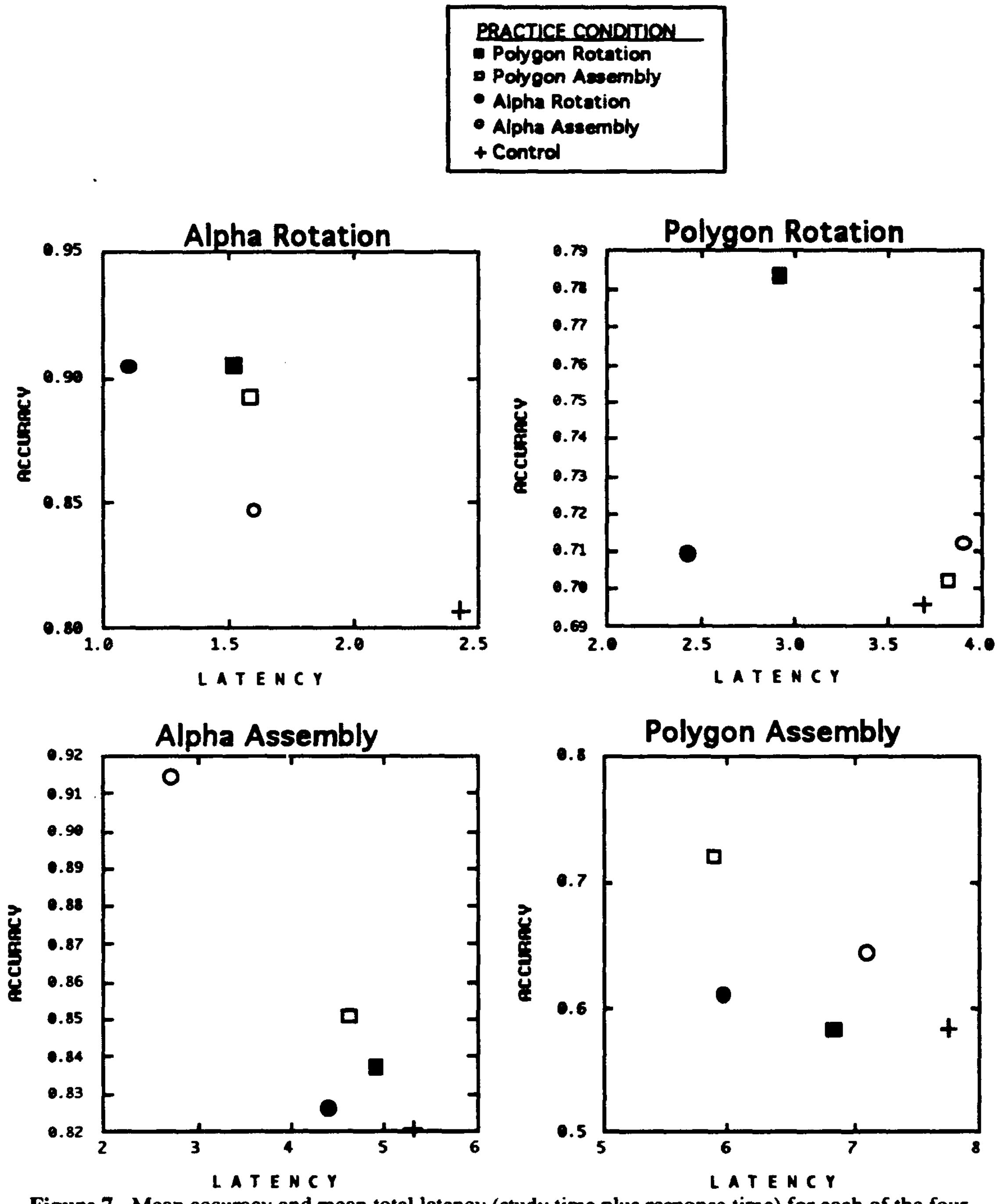


Figure 7. Mean accuracy and mean total latency (study time plus response time) for each of the four posttests, by practice condition (after Ackerman & Lohman, 1990).

Thus, data from self-paced posttest trials showed large positive transfer for repeating both transformation and stimulus set, moderate positive transfer for performing a practiced transformation on a new stimulus set, and slight negative transfer for performing a new transformation on an old stimulus set. However, different effects were sometimes observed for latencies and accuracies, suggesting that different conditions induced different speed-accuracy tradeoffs or that latency and accuracy can reflect different aspects of performance. Separating effects due to the learning of procedures or of stimulus sets from those due to speed-accuracy tradeoff requires that speed-accuracy tradeoff be controlled.

Results shown above were for 30 self-paced trials in each of the four conditions. Each subject was also administered a series of forced-paced trials in an effort to construct speed-accuracy curves. However, accuracy levels on forced-paced trials remained high even when stimuli were presented briefly. The problem was particularly severe for rotation of alphanumeric stimuli. Subjects may have been able to determine whether these stimuli were normal or mirror images without rotating them. Thus, successive presentation of problem stem and probe stimulus did not produce useful accuracy-latency tradeoff functions, especially for the alphanumeric stimuli.

Additional analyses of posttest data from this task showed several interactions between ability variables and treatments. Some interactions showed that treatments were disruptive for some subjects, particularly those who scored high on tests of verbal ability. Other interactions showed that order of tasks mattered, with first task attempted being particularly important. This suggests that a within-subjects design in which subjects attempt all four posttests is less appropriate than a between-subjects design in which each subject attempts only one posttest task.

In summary, although the analyses of data from self-paced trials showed several interesting effects, within-condition effects often differed for errors and for latencies, suggesting that a speed-accuracy analysis would be more appropriate. However, attempts to generate accuracy-latency curves were unsuccessful when problem stem and test probe were presented successively, particularly on alphanumeric stimuli. Further, interactions between ability and treatment variables suggested that the within-subjects design may have been inappropriate. The purpose of the first experiment in the present series was to revise this study to remedy these problems.

Experiment 1

The purpose of the first experiment was to determine whether practice on a particular type of spatial transformation (rotation or assembly) would produce transferable improvements in performing that same transformation on a different set of stimuli, or instead if practice would result in an improved representation for the practiced stimuli thus producing improved performance on trials requiring that a different transformation be performed on that same stimulus set. The first experiment was thus a direct extension of the Ackerman-Lohman study, with several modifications. First, problem stem and response probe were presented simultaneously rather than successively. Second, the contrast between alphanumeric and polygon stimuli was replaced with a contrast between two nominally parallel polygon stimulus sets. Third, the contrasts among posttest tasks were changed from within-subjects factors to between-subjects factors. Fourth, instead of attempting to adapt stimulus exposures for forced-paced trials to subjects, exposures were adapted to problems and the same set of stimulus exposures was used for all subjects.

Subjects (N=420 Air Force recruits) attempted 180 trials in which they practiced either assembling or rotating polygons. These practice trials were self paced, and feedback on response correctness (and, for the last 90 trials, on response latency) was provided after each trial and in summary form after each block of 30 trials. ISubjects were then administered several unrelated computer-based tasks. Finally, they attempted a 90-trial rotation or assembly task that contained either the same or different stimuli from those used in the practice task. Stimuli on the transfer task were presented for fixed exposures, and subjects required to respond within 750 msec of stimulus offset. Thus, the design contained 16 conditions defined by the crossing of pretest task (rotation or assembly), pretest stimulus set (A or B), posttest task (rotation or assembly), and posttest stimulus set (A or B).

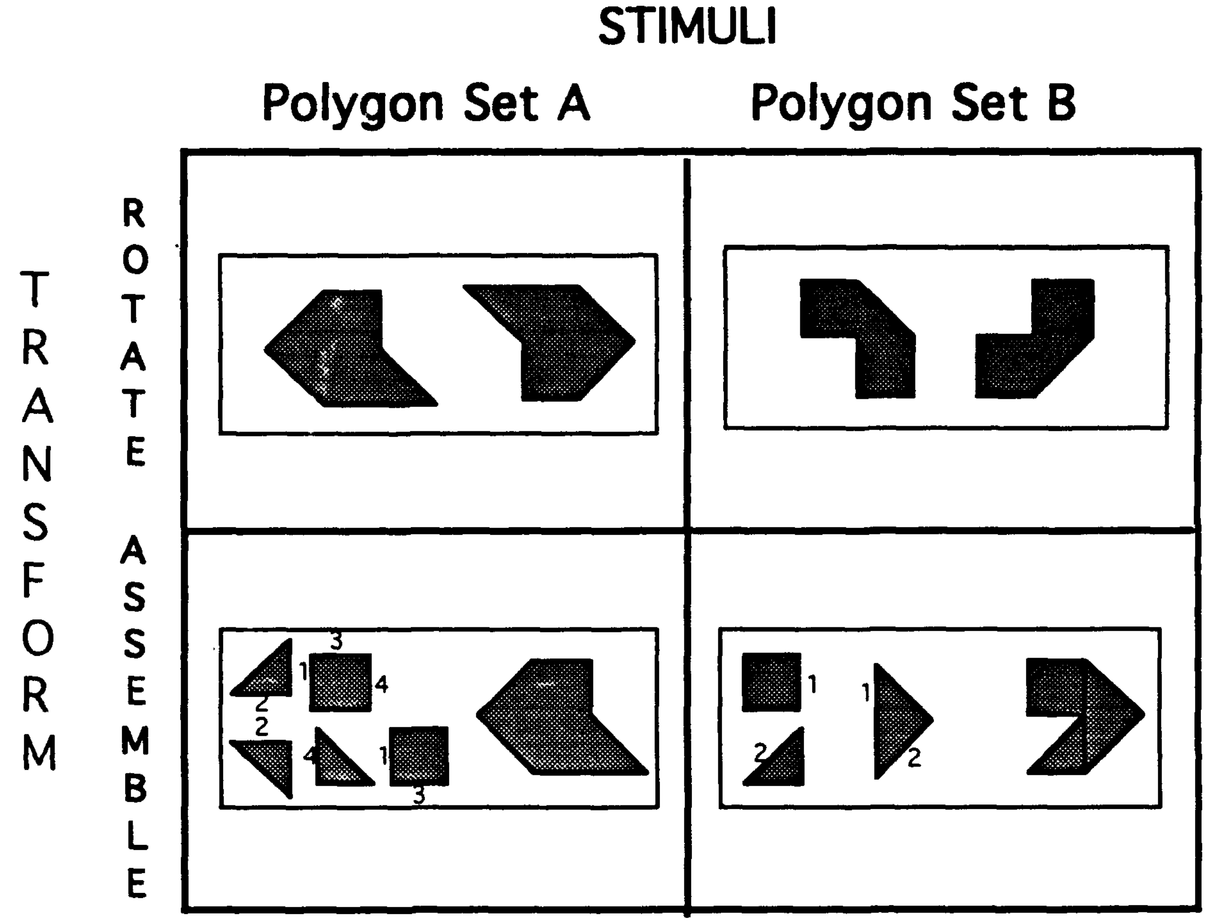


Figure 8. Design of pretest task and posttest task for Experiment 1. Design for the full experiment was obtained by crossing pretest task (4 conditions) with posttest task (4 conditions).

Materials and Procedure

Stimuli consisted of the 30 polygons shown in Appendix A. Sixteen of these forms were taken directly from the set developed by Alderton (1988). An additional six of Alderton's (1988) forms were modified to be nonsymmetric (and thus usable in the rotation trials). The remaining eight forms were new, although four had also been used in the Ackerman-Lohman (1990) study. Polygons were divided into two nominally parallel sets of 15 forms (based on ratings of form complexity and similarity), and one polygon from each pair was randomly assigned to Stimulus Set A and the other polygon to Stimulus Set B.

Assembly Trials. Each assembly trial contained a target figure on the right side of the screen and three, four, or five component figures on the left side of the screen. The component figures consisted of squares, triangles, and rectangles with color-coded sides indicating how they were to be combined to form the target figure. Sample assembly trials are shown in Figure 8 with number-coded rather than color-coded sides. Six trials were constructed for each of the 15 target figures in each stimulus set: one positive and one negative trial with three, four, or five component figures. Thus, each stimulus set was associated with 90 unique assembly trials.

Subjects who practiced assembling figures attempted one set of 90 trials twice, each time in a different random order. On each trial, subjects indicated whether the figure on the right of the screen would be created by combining the component figures on the left side of the screen in the manner indicated by pressing either the "l" (like) or "d" (different) key. For the first 90 trials, feedback on response accuracy was given after each trial; for the second 90 trials, feedback on both response accuracy and response latency was given after each trial.

The figure assembly posttest consisted of 90 fixed-exposure trials. Instruction screens and a demonstration problem are shown in Appendix B2. Subjects first practiced the procedure of responding as soon as the stimuli were replaced with pattern masks. These 10 trials were repeated until the subject responded within 750 ms of stimulus offset. Unlike previous studies, stimulus exposures were not estimated anew for each subject. Instead, a common set of 15 exposures (estimated in a pilot study) were used for all subjects. These exposures are shown in Table 1.

Table 1

Stimulus Exposures (in msec) for Fixed-Exposure Trials

	Stimulus Exposure 1	2	3	4	5
Assembly Task					
3-element	200	1650	3100	4550	6000
4-element	500	2500	4500	6500	8500
5-element	1500	3750	6000	8250	10500
Rotation Task					
45-pos	50	667	1300	1933	2550
90-pos	100	800	1500	2200	2900
135-pos	200	950	1700	2450	3200
all-neg	100	833	1550	2267	3000

Rotation Trials. Each rotation trial contained a standard figure on the left side of the screen and a rotated figure on the right side of the screen. Examples are shown in Figure 8 Instruction screens and a demonstration problem are shown in Appendix B1. Six trials were constructed for each of the 15 figures in each stimulus set: one positive and one negative trial at 45, 90, or 135 degrees of rotation. Foils were created by first reflecting the target figure about the horizontal axis. Thus, each stimulus set was associated with 90 unique rotation trials.

Subjects who practiced rotating figures attempted one set of 90 trials twice, each time in a different random order. On each trial, they pressed the "l" (like) or "d" (different) key to indicate whether the two figures could be rotated into congruence. For the first 90 trials, feedback on response accuracy was given after each trial; for the second 90 trials, feedback on both response accuracy and response latency was given after each trial.

The figure rotation posttest contained 90 fixed-exposure trials. Subjects first practiced the procedure of responding as soon as stimuli were replaced with pattern masks. These 10 trials were repeated until the subject responded within 750 ms of stimulus offset. Unlike previous studies, stimulus exposures were not estimated anew for each subject. Instead, a common set of 20 exposures estimated in a pilot study were used for all subjects. These exposures are shown in Table 1.

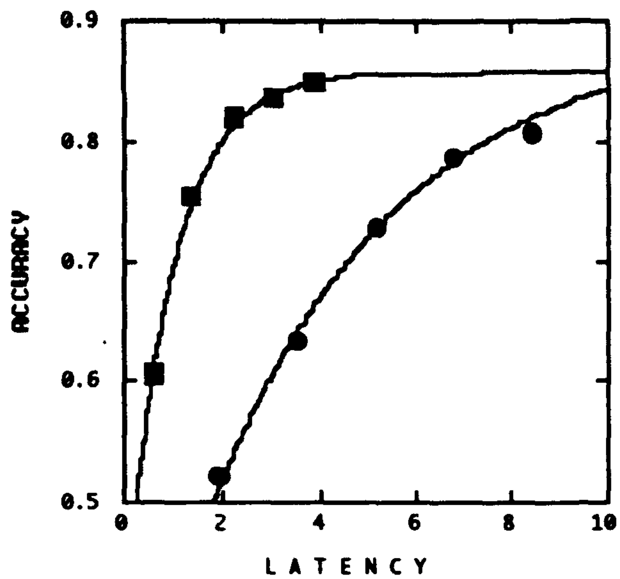


Figure 9. Probability correct as a function of total latency (exposure plus response latency) for the rotation (squares) and assembly (circles) posttests.

Results

Data averaged over subjects and stimulus sets are shown in Figure 9. Changes in the experiment from the procedures used in the previous study were successful in producing good accuracy-latency curves for both the assembly and rotation tasks. As expected, the assembly task was much more difficult than the rotation task. Performance was also more variable on the assembly task. Therefore, results are presented according to the nature of the posttest task.

Mental assembly task. A total of 281 subjects were administered the assembly task as posttest. We first eliminated 20 subjects who responded after the 750 ms deadline on 20 percent or more of the trials, and 50 subjects who failed to achieve an overall accuracy rate greater than 55 percent correct. Probability of a correct response and average total latency (exposure latency plus response latency) at each of the five exposure levels was then determined for each of the remaining 211 subjects.

The goal of the analysis was to estimate the effects of the experimental manipulations on each of the three parameters of a standard speed-accuracy model. This could be done by fitting a model to each subject's data, and then using the three model parameters for each subject as dependent variables in subsequent analyses. However, these parameters are often unstable, and usually show markedly skewed or truncated distributions. A better strategy is to fit an N-dimensional response surface to the data, after first screening subjects whose data do not conform to the general form of the model. Therefore, a nonlinear regression model of the form :

$$d' = \lambda\left(1 - e^{-\beta(t-\partial)}\right) \qquad (1)$$

was fitted to each subject's data, first using probability correct[5] and then d' as the dependent measure. d' was estimated by $d'_i = \text{probit}(z_i)\sqrt{2}$ (McNichol, 1972). Model fit statistics and parameter estimates for the two regressions were then compared. Results were indistinguishable for most subjects. However, models using d' as the dependent measure were less likely to converge and more likely to give unstable estimates, particularly for the curvature parameter. Discrepancies between the two models occurred most often for subjects who obtained unusually

[5] Probability correct was first adjusted by subtracting .5. As when d' is the dependent measure, the intercept then reflects the latency at which probability of a correct response is at the level of random responding.

low levels of accuracy on one of the longer exposure durations. Therefore, we eliminated subjects for whom the probability of a correct response was not significantly different from random responding for either of the two longest stimulus exposures (Condition 4 or 5 in Table 1)[6]. This left a sample of 188 subjects. We then examined the regression models and data for each subject and eliminated those subjects for whom the intercept was poorly estimated or for whom the overall model fit was poor ($R^2<.50$). This left a final sample of 149 subjects whose individual data well-conformed to the general speed-accuracy model. However, results were virtually identical to those obtained on the 188-subject sample. We report the results for the smaller sample here because plots of the data were a bit cleaner.[7]

Mean accuracy and mean total latency (presentation latency plus response latency) for each these 188 subjects at each of the five levels of stimulus exposure were then modeled using a generalization of Equation 1 in which the effects of practice task and practice stimulus set on the curvature, intercept, and asymptote parameters could be estimated (see Lohman, 1989):

$$P(C)-.5=\left(\lambda_0+\sum\lambda_i b_i\right)\left(1-e^{\left(\beta_0+\sum\beta_i b_i\right)\left(t-\left[\partial_0+\sum\partial_i b_i\right]\right)}\right) \qquad (2)$$

where P(C) is probability correct , t is total latency (exposure plus response latency), and λ_0, β_0, and ∂_0 are the average asymptote, curvature, and intercept parameters, respectively. The b_i coefficients define vectors that distinguish among particular contrasts. Three contrasts were represented in these models. The first contrast compared subjects who had practiced the same task (i.e., assembly, coded +1) with those who had practiced a different task (i.e. rotation, coded -1). The second contrast compared those who had practiced either rotating or assembling the same stimulus set (coded +1) with those who had practiced on a different stimulus set (coded -1). The third contrast represented the interaction between the first two contrasts. Results of the regression are shown in Table 2. Regression equations for each of the four conditions are shown in Table 3 and graphically in Figure 10. [8]

---

[6] The probability that average response accuracy for each subject differed from random responding was determined at all five exposure levels. If p<.80 at both exposure levels 1 and 2, then data at exposure level 1 were considered missing before individual models were fitted. This was necessary to avoid underestimating the intercept of the model. If p<.80 for three levels or at exposure level 4 or 5, then data were not modeled at the individual level.

[7] It is not uncommon to find that many recruits are either unwilling or unable to provide useful data, especially in the somewhat demanding and unusual forced-pace condition. Although we have yet to find any systematic difference between subjects who can and cannot adapt to the forced-pace procedure, subjects who fail to comply in the self-paced condition tend to be less able, and are probably also less motivated on average. This seems not to apply to other populations in other testing environments (only 13% of the data from student volunteers tested in our lab in Experiment 2 had to be discarded). But it also reflects a need for new and better ways to gather this sort of data.

[8] These models are derived from Table 2 by combining effects. For example, the curvature parameter for the "Same Task, Same Stimuli" condition is found by adding the effects for task (.030), stimuli (.031), and their interaction (-.038) to the average curvature (.257).

Table 2

Nonlinear Regression of Probability of a Correct Response on Total Latency, with Contrasts for Practice Task (Rotation or Assembly), Practice Stimulus Set (same or different), and their Interaction, for the Assembly Posttest (N = 149)

Parameter	Estimate	SE
Curvature (ß)	.257	.042
Task	.030	
Stimuli	.031	
Task x Stimuli	-.038	
Intercept (∂)	1.800	.105
Task	-.165	
Stimuli	-.066	
Task x Stimuli	.017	
Asymptote (L)	.393	.038
Task	.012	
Stimuli	-.030	
Task x Stimuli	.028	

Note. $R^2 = .57$ (df = 12 for model, 703 residual)

Table 3

Curvature (ß), Intercept (∂), and Asymptote (λ) Parameters for the Regression of Probability Correct on Total Latency, and Probability Correct at the Longest Stimulus Exposure for each of the Four Combinations of Practice Task (Assembly or Rotation) and Stimulus Set (A or B) for the Assembly Posttest (N = 149)

Condition Task	Stimuli	ß	∂	λ	Acc(5)[a]
Same	Same	.280	1.59	.403	.836
Same	Different	.294	1.68	.408	.846
Different	Same	.295	1.88	.322	.779
Different	Different	.158	2.05	.439	.767

[a] Probability correct at the fifth or longest stimulus exposure. This score is one of the dependent measures, not one of the model parameters. Compare with $\lambda + .5$.

There are several noteworthy effects here. Recall that the Cascade model predicts that familiar stimuli will show higher asymptotic levels of activation than will unfamiliar stimuli. However, Figure 10 shows that this did not occur. Instead, curves for the two conditions in which subjects practiced the same assembly task (squares in Figure 10) had much higher asymptotes than the two conditions in which subjects practiced the rotation task (circles in Figure 10). Subjects who practiced the same task (i.e., assembly) also showed higher levels of accuracy at intermediate exposures as well, suggesting more efficient information processing (Thomas, 1974). One explanation for this result is that practice allows subjects to reduce attentional resources allocated to the assembly transformation, thereby increasing resources available for short-term storage (Shiffrin & Schneider, 1977). Perhaps subjects developed or further automatized a general purpose mental assembly skill. This seems unlikely, since practice was relatively brief and other evidence suggests that such skills are acquired during childhood (Piaget & Inhelder, 1967). More likely, subjects who practiced the assembly task in the less demanding self-paced condition (where feedback was also provided) learned how to perforr the task more efficiently. In particular, they learned what sorts of assembly operations were required in the task. This may have involved learning to attend to particular aspects of problems that initially seemed irrelevant, and to ignore other aspects of the problems that initially seemed relevant. Or it may simply have allowed subjects time to develop more efficient strategies for selecting and combining elements. Reducing the problem space in this way would have several benefits. First, it would allow subjects to achieve a correct answer sooner. Inspection of Figure 10 shows that this was indeed the case: the two "same task" conditions obtained higher levels of accuracy at short and intermediate exposures than did the two "different task" conditions. Better performance at the shortest exposure is reflected in a significantly smaller intercept for the same task condition ($\partial = 1.63$ vs. 1.97 for same vs. different task, respectively; **t**=3.40, **p**<.01). Second, reducing the problem space means that practiced subjects would be less likely to be misled by extraneous elements in problems, which not only take time but also increase the likelihood of false alarms on negative trials. Third, reducing the problem space means that subjects would be able to devote more working memory resources to maintaining an image of the to-be-assembled figure. The result would be a higher level of accuracy at all exposure levels but particularly on long exposure trials, which is the most striking feature of the data in Figure 10. This effect is not well represented by the asymptotes of the models, since the asymptotes estimate accuracy when infinite time is allowed. Inspection of Figure 10 shows that small changes within the response space can lead to dramatic differences outside of it. Therefore, we analyzed accuracy scores on trials with longest exposure (the accuracy analog to the intercept). This score is labeled Acc(5) in Table 3. The effect for task was highly significant (**F**(1,145)= 25.63, **p**<.0001).

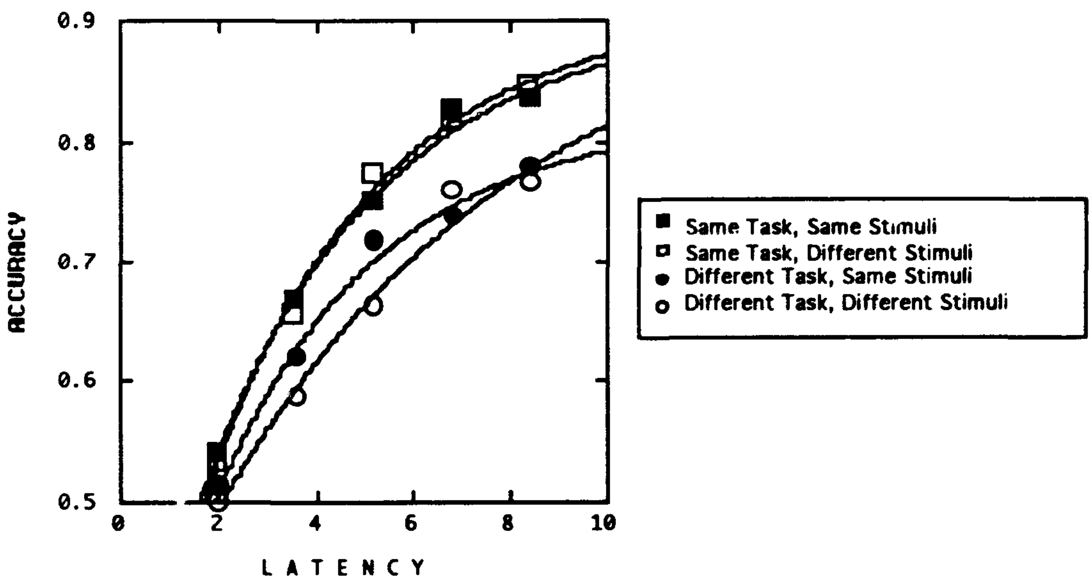


Figure 10. Accuracy-latency curves for the four practice conditions on the assembly posttest (N=149).

What about stimulus effects? The Cascade model predicts higher levels of asymptotic accuracy for practiced stimuli. On the one hand, when subjects practiced and were posttested on the assembly task (squares in Figure 10), then whether stimuli had also been practiced (filled symbols in Figure 10) or had not been practiced (open symbols in Figure 10) did not matter. Means for these two conditions generally overlapped. On the other hand, if subjects had practiced the rotation task (but were tested on the assembly task), then having practiced on the same stimuli conferred some advantage. Curvature parameters for these two conditions differed markedly (β = .295 vs. .158, **t**=3.42, **p**<.01). Thus, contrary to the predictions of the Cascade model, practicing a particular set of stimuli affected the curvature parameter rather than the asymptote parameter, but only if subjects had practiced rotating the stimuli rather than assembling them.[9]

Figure 10 shows that subjects who practiced the assembly task achieved higher levels of accuracy than subjects who practiced the rotation task. But how large were these effects? Figure 11 shows plots of effect sizes using the different task, different stimuli condition as the standard.[10] The same symbols used in Figure 10 were also used in Figure 11 to facilitate comparison. For accuracy, effect sizes ranged from .5 SD at the shortest exposure level to 1.2 SD at the middle exposure level for subjects who practiced the same (i.e., assembly) task in the practice phase. The average effect size was approximately .82 SD in both same-task conditions (squares in the left panel of Figure 11). These are large effects.

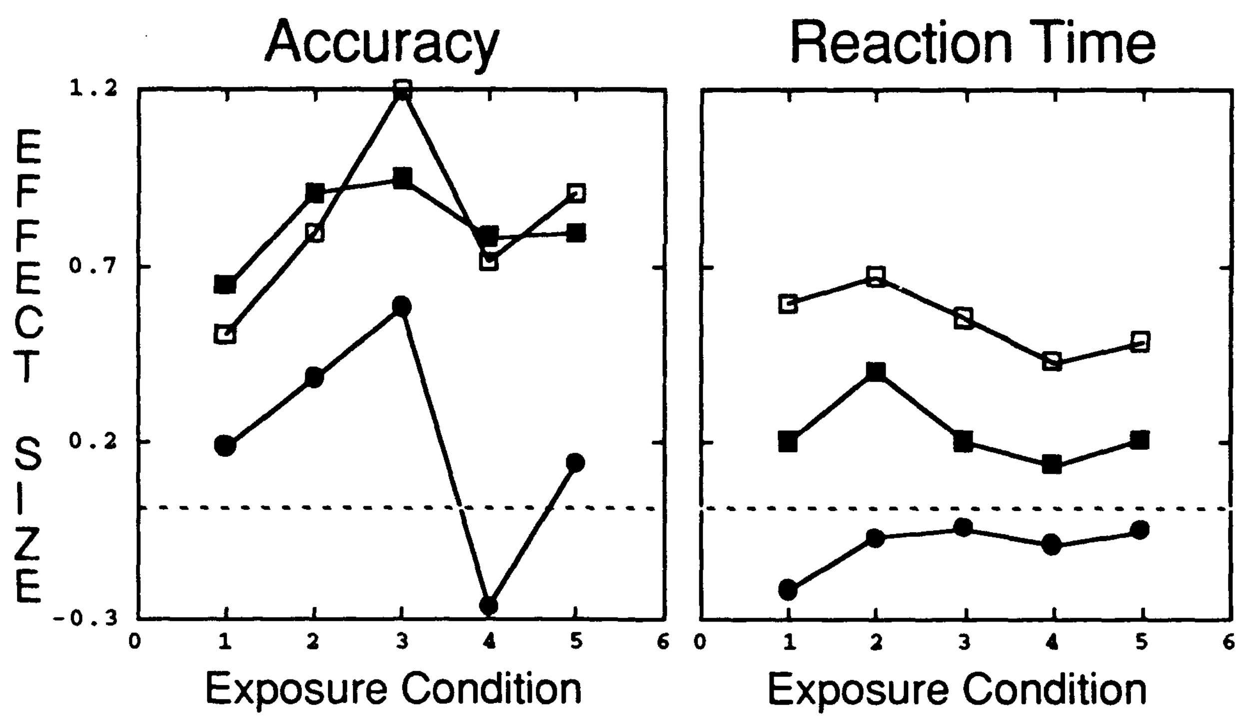


Figure 11. Effect sizes for the assembly posttest using the different task, different stimuli condition as the control group. Squares are same task; filled symbols are same stimuli.

---

[9] This clarifies the finding in the Ackerman-Lohman (1990)study that practice in rotating polygons resulted in faster but not more accurate performance on the polygon assembly posttest (see Figure 6).

[10] Effect sizes were computed by subtracting mean accuracy at a given exposure level for the different task, different stimuli condition from the corresponding means of the other three conditions and then dividing by the SD for the different task, different stimuli condition. Subtraction of means was performed in the reverse order for reaction times, thus making positive effects indicative of better performance for both accuracy and reaction time.

In contrast, practicing with the same stimulus set but on a different task (i.e., rotation) had much smaller effects (filled circles in the left panel of Figure 11). These effect sizes ranged from -.26 to .58 SD with an average value of .20 SD. [The negative effect for exposure level 4 is clearly anomalous in all plots.] This small advantage for practicing the same stimulus set did not occur in the same-task condition (open vrs filled squares in the left panel of Figure 11). Thus, subjects learned something about stimuli when rotating them that they did not learn when assembling them.

Exposure latency is constrained to be equal across the four conditions, thereby moving most of the variation among conditions to the accuracy score. However, the total latency score plotted on the abscissa in Figure 10 is the sum of exposure latency and reaction time once the stimulus was removed from view. Exposure latencies are much larger than reaction times, and thus dominate the overall accuracy-latency plots. Nevertheless, there were interesting and significant differences in reaction times among the four conditions. Effect sizes for reaction times are plotted in the right panel of Figure 11. The largest differences among conditions occurred at the shortest exposure (these correspond to intercept differences in Figure 10). Average effect sizes were .55, .23, and -.10 SD for the same task, different stimuli, same task-same stimuli, and different task-same stimuli conditions, respectively. In absolute units, the differences were 46, 19, and -8 ms, respectively. Thus, subjects who had practiced the assembly task on different stimuli were on average 27 ms faster than subjects who had practiced the assembly task on the same stimuli. Perhaps subjects who practiced assembling stimuli learned to remember them as an assemblage of simpler shapes rather than as a unit. This would increase comparison latencies (Cooper, 1982; Lohman, 1988)..

In summary, for the assembly posttest, having practiced the assembly task led to a significant reduction in intercept, more accurate performance on intermediate exposure trials, and an even larger increase in final level of accuracy achieved on long exposure trials. The first effect suggests that task-specific procedures were learned, the second effect suggests more efficient information processing, and the third effect suggests a significant increase in the amount of information that could be maintained in working memory. McClelland's (1979) model led us to predict that asymptotic levels of performance would be higher for practiced than for nonpracticed stimuli, an effect we and others have observed in priming studies. However, the fact that stimulus set had no effect on final accuracy levels suggests that these effects could not be attributed to stronger memory traces for practiced stimuli. Also unexpected was the fact that stimulus set had the only significant effect on the curvature parameter and then only when subjects had practiced rotating stimuli, not when they had practiced assembling them. According to McClelland's model, changes in curvature should reflect changes in a relatively slow or rate-limiting process. Finally, an examination of effect sizes showed that these practice-induced changes were not only statistically significant, but practically significant as well.

<u>Mental rotation task</u>. A total of 302 subjects were administered the rotation task as posttest. We first eliminated subjects who responded after the 750 ms deadline on 12 percent or more of the trials, and those who failed to achieve an overall accuracy rate greater than 55 percent. Probability of a correct response at each of the five exposure levels was then determined for each of the remaining 248 subjects. Once again, we eliminated seven subjects for whom the probability of a correct response was less than .80 for exposure condition 4 or 5. Nonlinear regression models of the form shown in Equation 1 were then fitted to each subject's data using probability correct as the dependent measure. Subjects who showed poor model fit (as judged by RMSE > .06) were then deleted. The final sample thus consisted of 205 subjects.

Mean accuracy and mean total latency (presentation latency plus response latency) at each of the five levels of stimulus exposure were then modeled using Equation 2. Contrasts for task, stimulus set, and their interaction were coded as in the previous analysis. The results are shown in Table 4. Regression models for the four conditions are shown in Table 5 and then

Table 4

Nonlinear Regression of Probability of a Correct Response on Total Latency, with Contrasts for Practice Task (Rotation or Assembly), Practice Stimulus Set (same or different), and their Interaction, for the Rotation Posttest (N = 205)

Parameter	Estimate	SE
Curvature (ß)	1.011	.109
Task	.062	
Stimuli	.063	
Task x Stimuli	.061	
Intercept (∂)	.155	.057
Task	.002	
Stimuli	-.007	
Task x Stimuli	.016	
Asymptote (L)	.354	.007
Task	.036	
Stimuli	.005	
Task x Stimuli	.001	

Note. R^2 = .48 (df = 12 for model, 1005 residual).

Table 5

Curvature (ß), Intercept (∂), and Asymptote (λ) Parameters for Within-Condition Regressions of Probability Correct on Total Latency, and Probability Correct at the Longest Stimulus Exposure for each of the Four Combinations of Practice Task (Assembly or Rotation) and Stimulus Set (A or B) for the Rotation Posttest (N = 205)

Condition Task	Stimuli	ß	∂	λ	Acc(5)[a]
Same	Same	1.197	.166	.396	.887
Same	Different	.949	.148	.384	.870
Different	Same	.951	.130	.322	.815
Different	Different	.948	.176	.314	.811

[a] Probability correct at the fifth or longest stimulus exposure. This score is one of the dependent measures, not one of the model parameters. It should be comparable to λ + .5. Effect for task $F(1,200) = 32.06$, $p < .001$.

graphically in Figure 11. The results were straightforward. Task similarity had large and significant effects on final accuracy levels achieved ($F(1,200)=32.06$, $p<.01$), and on asymptote parameters ($t=5.14$, $p<.01$). There was a small effect for practicing the same stimulus set used on the posttest, but only when the task was also the same. As on the assembly posttest, practice in assembling stimuli did not improve performance. Practice in rotating them did. Also as on the assembly task, the effect unexpectedly emerged in the curvature parameters (β = 1.197 and .949 for the rotation task with same and different stimuli, respectively). Differences between intercepts were small and nonsignificant.

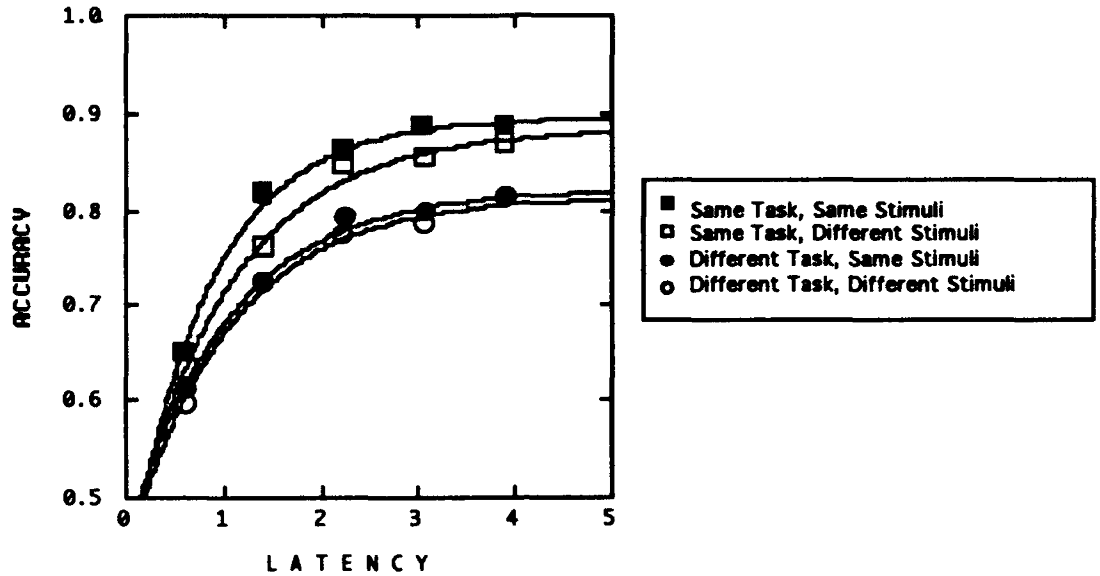


Figure 12. Accuracy-latency curves for the four practice conditions on the rotation posttest (N=205).

Figure 12 shows that subjects who practiced the rotation task achieved higher levels of accuracy on the rotation posttest than subjects who practiced the assembly task. But how large were these effects? Figure 13 shows plots of effect sizes using the different task, different stimuli condition as the standard. As before, the same symbols used in Figure 12 were also used in Figure 13 to facilitate comparison. For accuracy, effect sizes ranged from approximately .4 SD at exposure levels 1 and 2 to almost 1.2 SD at exposure level 4 for subjects who practiced the same (i.e., rotation) task in the practice phase. The average effect sizes were .60 SD and .91 SD for the same task, different stimuli (open squares) and same task, same stimuli conditions (filled squares), respectively. These effects are roughly the same size as those observed on the assembly posttest.

In contrast, practicing with the same stimulus set but on a different task (i.e., assembly) had much smaller effects. These ranged from -.01 to .20 SD with an average value of .11 SD. Comparing effect sizes for the two same-task conditions (i.e., the top two lines in the left panel of Figure 13) gives a somewhat larger estimate of .30 SD for the benefit of practicing the same stimulus set used in the posttest. Thus, unlike the previous analysis on the assembly task, practicing on the same stimulus set generally conferred some advantage, especially when the task was also the same.

Once again, effect sizes for reaction times showed interesting results. These are plotted in the right panel of Figure 13. In contrast to the assembly posttest, differences among conditions were smallest at the shortest exposure and then increased. Average effect sizes were .18, .00, and -.12 SD for the same task, same stimuli, same task-different stimuli, and different task-same stimuli conditions, respectively. In absolute units, the differences were 13, 0, and -8 ms, respectively. Although small, the effects are consistent with the hypothesis that practice in assembling particular stimuli results in slightly slower reaction times when mentally rotating those same stimuli.

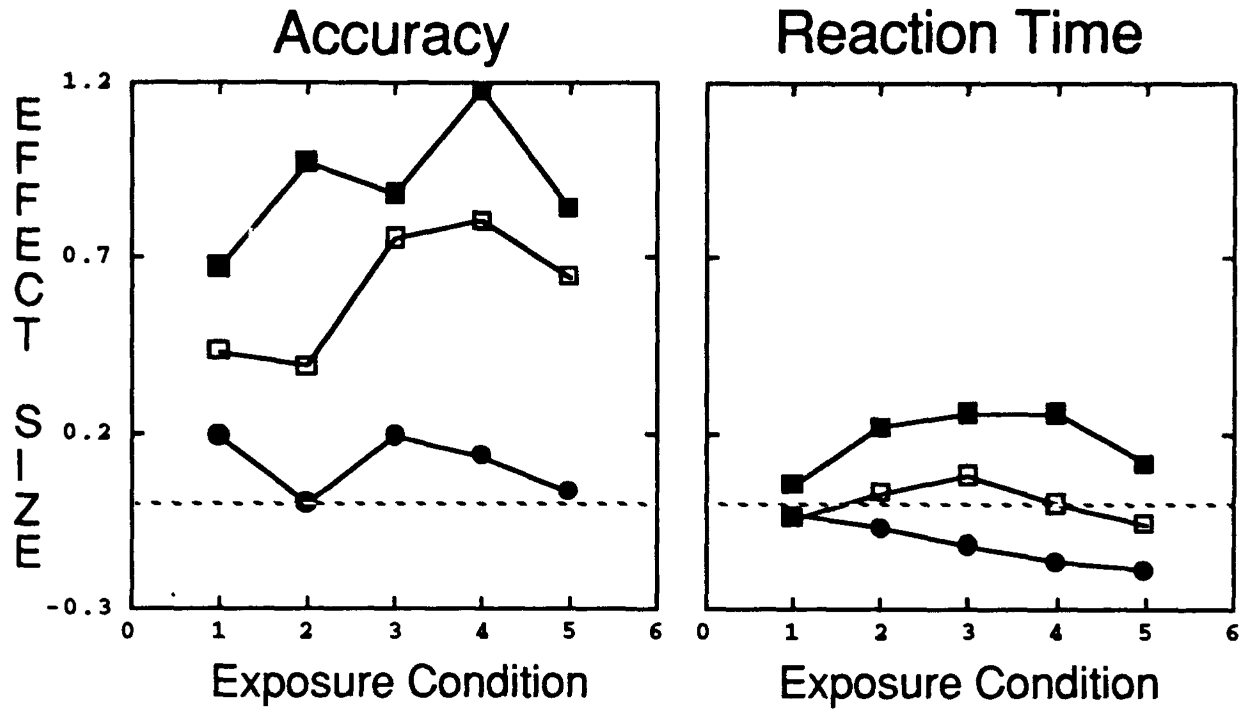


Figure 13. Effect sizes for the rotation posttest using the different task, different stimuli condition as the control group. Squares are same task; circles are same stimuli.

Discussion

For both the assembly and rotation posttests, practicing the same task (i.e., assembly or rotation) in the self-paced session resulted in much higher levels of accuracy on long-exposure trials, a more rapid rise to this asymptote, and, at least in the case of the assembly task, an earlier rise above the level of random responding. Thus, procedural transfer was the primary effect of practice. Stimulus transfer was smaller and more specific. Practice in rotating stimuli lead to more efficient processing of those stimuli on both the assembly and rotation posttests. However, such practice did not result in higher levels of accuracy on relatively unspeeded trials. Several factors may have caused the effect for stimulus set to be smaller than expected, to appear in the curvature rather than in the asymptote, and to occur only when subjects practiced rotating stimuli and not when they practiced assembling them. First, stimuli were all regular polygons and thus were probably all relatively familiar to subjects. Practice effects may have been larger had we used unfamiliar shapes. Second, practice on truly unfamiliar shapes may have resulted in significant increases in asymptotes, as the Cascade model predicts. However. practice-induced changes in the representation of familiar shapes were more subtle. Attempting to rotate stimuli may have induced subjects to represent stimuli holistically. This would increase the rate at which such stimuli could be rotated or otherwise transformed but would not produce much change in their familiarity. Practice in assembling stimuli from various component pieces had no such effects, probably because subjects may have focused on the component shapes and attended only incidentally to the target shape. Further, such practice may have encouraged subjects to maintain a representation of the overall shape as an assemblage of simpler shapes. The small increase in reaction time on the rotation posttest for subjects who had practiced assembling figures is consistent with this hypothesis. Also consistent was the finding that reaction times on the assembly posttest were significantly longer when subjects practiced assembling the same stimuli used in the posttest than when subjects practiced assembling different stimuli. Thus, there was some indication here of the interference effects observed in the Ackerman-Lohman study (see Figure 5). It is of some interest that these effects emerged only in the analyses of comparison latencies. Previous work has shown this measure to be particularly sensitive to the nature of the representation a subject has established (Cooper, 1982). Finally, effects due to stimuli may have been small because practice had different effects on different subjects. For example, Bethell-Fox and Shepard (1988) found that, even after extensive practice, some subjects represented stimuli in a manner that made amount of rotation dependent on the complexity of the stimulus. Other subjects showed effects of stimulus complexity only early in rotation practice.[11] .

In summary, the primary product of practice was the learning of task procedures. Effects for procedural transfer were large for both the assembly and rotation tasks. However, the study had several important limitations. First, practice tasks and posttests differed only in the pacing of items, thus maximizing specific transfer. We wondered how much procedural transfer would be observed if subjects practiced somewhat different tasks in a medium (paper-and-pencil) that differed from the medium of the computer-based posttest. Second, both this experiment and the earlier Ackerman-Lohman study provided only a relatively small amount of practice. This was required in the present experiment because recruits were available only for a single three-hour session. We wondered if stronger effects for stimuli might be observed if subjects were given more practice. Third, some have argued that the acquisition of general procedural knowledge requires either practice in a variety of formats or practice with concrete materials. If this is true, then it may be necessary to move outside of the computer lab to develop more general spatial skills. The second experiment was designed to investigate these questions.

Experiment 2

The second study addressed the issue of whether different types of practice affect different aspects of spatial skill acquisition. Subjects (N=144 university students) practiced assembling shapes in different media: plastic shapes (Tangoes), paper-and-pencil drawings (Paper Form Board), or a computer game (Tetris). Other subjects received practice in all three media. Practice tasks also varied in the similarity of their procedures and stimuli to those used on the posttest. Four questions about the effects of practice on the improvement and transfer of spatial abilities were investigated.

First, we wondered how much of the transfer observed in the previous experiment could be attributed to identity of procedures. Would changing the format of the practice task from computer-based to paper-and-pencil reduce transfer to a computer-based posttest? What would happen if the practice task were made similar but not identical to the transfer task? In particular, suppose problems sometimes required the subject to perform procedures not required on the posttest. Would these varied task demands prevent subjects from assembling and tuning a strategy that could be transferred to the posttest? In an effort to answer these questions we developed a condition (Form Board) in which subjects practiced a task that was similar to the posttest, but that was administered in a different medium (paper-and-pencil vrs. computer). Subjects practiced at least two different types of problems in each of three practice sessions. On all problems, subjects were required to number the sides of five component figures to indicate how they should be combined to form a target figure. However, some problem sets contained a sixth component figure that was not needed. On other sets, one of the five component figures had to be rotated to a new orientation before it could be used. In this way we hoped to vary task procedures so that subjects did not simply assemble and proceduralize a strategy that could be directly transferred to the posttest since none of the items on the posttest contained extra or rotated component figures. Paper-and-pencil problems differed from computer-administered

[11] However, Ackerman and Lohman (1990, Experiment 1) found that high verbal subjects who practiced assembling polygons were more accurate when rotating polygons than were high verbal subjects who practiced assembling alphanumerics.

problems in other ways as well. Subjects were free to work on practice problems for as long as they wished whereas posttest problems were presented for fixed, mostly brief exposures. On practice problems, subjects had to determine how component figures could be combined and number them accordingly. On posttest problems, sides of component figures were color-coded to indicate how they should be combined. Subjects merely accepted or rejected the target figure. On practice problems, a single target figure was shown at the top of the page and applied to all problems on that page. On test problems, the target figure appeared to the right of each set of component figures and did not necessarily show how component figures should be combined. Finally, component figures used in the practice problems were generally more irregular and thus more difficult to remember and combine than the component figures used on the posttest. Thus, performance of subjects in the Form Board condition allows inference to the case in which examinees have practiced problems in a paper-and-pencil format that are nominally similar to those presented on a computer-administered test.[12]

The second question concerned how much of the transfer observed in previous studies could be attributed to practice in taking a computer-based test under time pressure. There are really two issues here. The first is simply how much benefit accrues from practice in using the computer to take a test. We expected at best small benefits for this computer-literate population. All had presumably already developed general procedures for interacting with computers (although subjects might still be expected to show some evidence of further proceduralization of the specific skills required). The larger issue was whether practice in responding quickly on a spatial task in which item exposure was not under the subject's control would transfer to the posttest. In particular, we wondered whether subjects who had practiced playing a computer game that in a general way resembled the posttest would have an advantage on the fixed-exposure posttest. For this purpose, we selected a computer game called Tetris. In this game, shapes constructed from three adjacent squares of different colors fall from the top of the screen at a preset rate. The shape can be moved left or right, and colors of squares shifted before the shape makes contact with squares that have accumulated on the bottom of the screen. The goal is to create horizontal, vertical, or diagonal rows of three or more squares of the same color (which then disappear).

The third question once again addressed the issue of stimulus familiarity. We wondered whether the failure in Experiment 1 to obtain an effect for practice in assembling target stimuli used on the posttest could be attributed to the relative brevity of practice. Thus, 15 of the 30 target stimuli used in the posttest were also used as target stimuli on practice exercises for subjects in the Form Board condition. Subjects practiced assembling these stimuli for three one-hour sessions on consecutive days rather than for the single one-hour session used in Experiment 1.

Finally, the fourth question concerned whether we might be able to induce more general improvements in spatial assembly skill, improvements that would show broader transfer. There are two rather different views about how this may occur. On the one hand, theories of ability development based on internalization of an external activity (e.g., Piaget & Inhelder, 1967) argue that children first learn to operate on concrete objects before they can operate on mental representations of those objects. These theories predict that practice in assembling concrete objects would be more likely to produce a general improvement in the ability to assemble forms than would practice on paper-and-pencil drawings, or even varied practice in different contexts

---

[12] Each of these factors (and others) could have been studied by creating conditions in which they varied systematically. (Indeed, some were tested in other conditions in the experiment). However, we did not have the resources (or the inclination) to do so. Our thinking was that if transfer was not observed or was substantially reduced from the level obtained in Experiment 1, then we could later determine why. If transfer was observed, then such concerns would be moot.

and formats. Accordingly, we had subjects in one condition practice a game called Tangoes. In this game, two players compete to be the first to construct a target picture. Each player has a set of seven plastic forms: five triangles, one square, and one parallelogram, all of which must be used to construct the target. Target forms are sometimes regular polygons (as in the example problem in Appendix D), but more often are pictures of common objects (such as a sailboat). The main similarity between the game and the posttest is that both require the assembly of a reasonably well-formed target figure using simple geometric shapes. Thus, if transfer was observed, it would be of a fairly general sort. On the other hand, theories of skill acquisition based on stimulus generalization (e.g., Anderson, 1983) suggest that the acquisition of general skills requires practice of that skill in varied contexts. Therefore, subjects in the Mixed condition practiced all three assembly tasks: the paper-and-pencil Form Board exercises, the Tangoe game with its plastic shapes, and the computer-based Tetris game.

Predictions from the Cascade Model

McClelland's (1979) model suggests that improvement in a general, rate-limiting transformation will be reflected in the curvature of an accuracy-latency function (see Figure 2.) Accordingly, we expected that subjects who practiced all three tasks might show a steeper curvature on the posttest than subjects in other conditions. Alternatively, if internalization theories of ability development are correct, then subjects in the Tangoes condition should show the steeper curvature. Second, we expected that subjects who practiced the computer game (Tetris) to automatize some of the specific skills required for interacting with the computer. McClelland's (1979) model suggests that changes in fast processes will be reflected in the intercept of an accuracy-latency function. Thus, we hypothesized that the accuracy-latency curve for these subjects would have a lower intercept. Third, McClelland's model suggests that familiar stimuli will achieve higher levels of asymptotic activation than unfamiliar stimuli. Therefore, we expected that subjects in the paper-and-pencil condition who practiced on specific stimuli used in the posttest would show a higher asymptote for these stimuli than for nonpracticed stimuli. Experiment 1 showed that subjects who practiced procedures used on the posttest also achieved higher asymptotic levels of accuracy. We therefore expected that the accuracy-latency curve for subjects in the Form Board condition would have a higher asymptote than the accuracy-latency curves for subjects in other conditions. But we also expected that the effect would not be as large as in the previous experiment in which practice and posttest tasks were the same.

Method

Subjects were randomly assigned to one of four practice conditions or to a no-practice control. Each condition contained an equal number of male and female undergraduates. Subjects practiced assembling polygons for three, one-hour sessions on consecutive days. Subjects in Group 1 (Paper Form Board) practiced on paper-and-pencil form board problems. Example exercises are shown in Appendix C. Target stimuli in these exercises constituted half of the target stimuli used in the posttest, although the problems differed (see below). Subjects in Group 2 (Tangoes) practiced a puzzle task called Tangoes in which two players compete to be the first to construct a target picture. Each player has a set of seven plastic forms: five triangles, one square, and one parallelogram, all of which must be used to construct the target. An example problem is shown in Appendix D. Subjects in Group 3 (Tetris) practiced a computer game called Tetris in which shapes constructed from three adjacent squares of different colors fall from the top of the screen and must be rotated or shifted before they contact squares that have accumulated on the bottom of the screen. Subjects in Group 4 (Mixed) practiced all three tasks (Form Board, Tangoes, Tetris) for one session each.

During the fourth session (or first session for control subjects), subjects were administered a computer-based posttest in which 180 figural assembly problems were presented for the same set of fixed exposures used in the previous experiment (see Table 1). The goal of this testing procedure was to generate a speed-accuracy curve for each subject. Thus, the most rapid exposure was designed to elicit responses whose correctness was at the level of random guessing; the longest exposure was designed to elicit the subject's most accurate performance when using the same strategy employed on more rapidly paced trials. Subjects were first given ten trials on which they practiced responding as soon as possible after stimulus offset. These trials were repeated until subjects could respond within 750 msec. A sample item is shown in Appendix B2.

Trials on the posttest were constructed by decomposing each of 30 geometric target figures into three, four, or five component shapes. Component shapes consisted of squares, triangles, and rectangles, the sides of which were color coded to indicate how they should be assembled. Component shapes were shown on the left side of the screen and the target figure was shown simultaneously on the right side of the screen. The 30 target figures shown in Appendix A were divided into two sets of 15 stimuli, called Set X and Set Y. Subjects in the Form Board condition practiced assembling stimulus set A, but from less regular and hence more difficult component shapes than those used in the posttest.[13]

Results

As in the previous experiment, we first eliminated subjects who failed to solve more than 55 percent of the trials, or who were unable to execute a response within 750 ms after stimulus offset. This reduced the sample size from 144 to 125, with approximately the same number of subjects in each of the five experimental conditions.

Mean accuracy and mean total latency (presentation latency plus response latency) at each of the five levels of stimulus exposure were then modeled using a generalization of Equation 1 in which the effects of practice task and practice stimulus set on the curvature, intercept, and asymptote parameters could be estimated (see Lohman, 1989):

$$P(C)-.5=\left(\lambda_0+\sum\lambda_i b_i\right)\left(1-e^{\left(\beta_0+\sum\beta_i b_i\right)\left(t-\left[\partial_0+\sum\partial_i b_i\right]\right)}\right)$$

where P(C) is probability correct, t is total latency (presentation plus response), and λ_0, β_0, and ∂_0 represent the average asymptote, curvature, and intercept parameters, respectively. The b_i coefficients define vectors that distinguish among particular contrasts. Seven contrasts were coded. The first contrast compared stimulus sets X (coded +1) and Y (coded -1). The next four contrasts compared subjects who had practiced a particular task (e.g., Tangoes, coded +1) with the control group (coded -1). The sixth contrast represented the interaction between stimulus set and Form Board, and the seventh contrast between stimulus set and Mixed group. These were the only groups that had practiced stimulus Set X. The interaction terms thus tests the hypothesis that performance of the particular group differed from that of the control group on practiced vs. nonpracticed stimuli. The results are shown in Table 6. Clearly, the interaction contrasts were small and nonsignificant for all three parameters. Therefore, the contrasts for stimulus set and its interactions were dropped, and the model fitted again. Regression models shown in Table 7 (and graphically in Figure 12) are based on these reduced models.

[13] We intended to use the nominally parallel stimulus Sets A and B used in the previous experiment. However, due to a change in the computer program, a different set of 15 stimuli (called Set X) were used. The remaining 15 stimuli (Set Y) appeared to be slightly more difficult (see Table 6).

Table 6

Nonlinear Regression of Probability of a Correct Response on Total Latency, with Contrasts for Practice Task (Form Board, Tangoes, Tetris, or Mixed), Practice Stimulus Set (A or B), and the Interaction of Form Board Practice and Stimulus Set, and Mixed Practice and Stimulus Set on the Assembly Posttest (N = 125)

Effect	Parameter ß	∂	λ
0 Control	.2384	1.5809	.3494
1 Stim Set	-.0128	-.0676	.0267
2 Tangoes	.1421	.2611	-.0939
3 Tetris	-.0633	-.2777	.0286
4 Form Board	-.0272	-.1533	.0619
5 Mixed	-.0017	.0778	-.0018
6 Stim x FB	.0015	.03710	-.0021
7 Stim X Mx	.0042	.0507	.0082
Standard Error	.0435	.1258	.0396

Note. $R^2 = .37$ (df = 24 for model, 1226 residual)

Table 7

Curvature (ß), Intercept (∂), and Asymptote (λ) Parameters for Within-Condition Regressions of Probability Correct on Total Latency, and Probability Correct at the Longest Stimulus Exposure for each of the Five Experimental Groups

Condition	ß	∂	λ	Acc(5)[a]
Control	.2130	1.5455	.3763	.744
Tangoes	.3468	1.8128	.2650	.722
Tetris	.1255	1.1018	.4526	.775
Form Board	.2189	1.4967	.4075	.811
Mixed	.2175	1.6541	.3653	.774
	.0474	.1602	.0655	

Note. Parameters from a reduced model in which stimulus set and interactions were omitted.

[a]Probability correct at the fifth (longest) exposure. These are dependent measures, not parameters of the regression model.

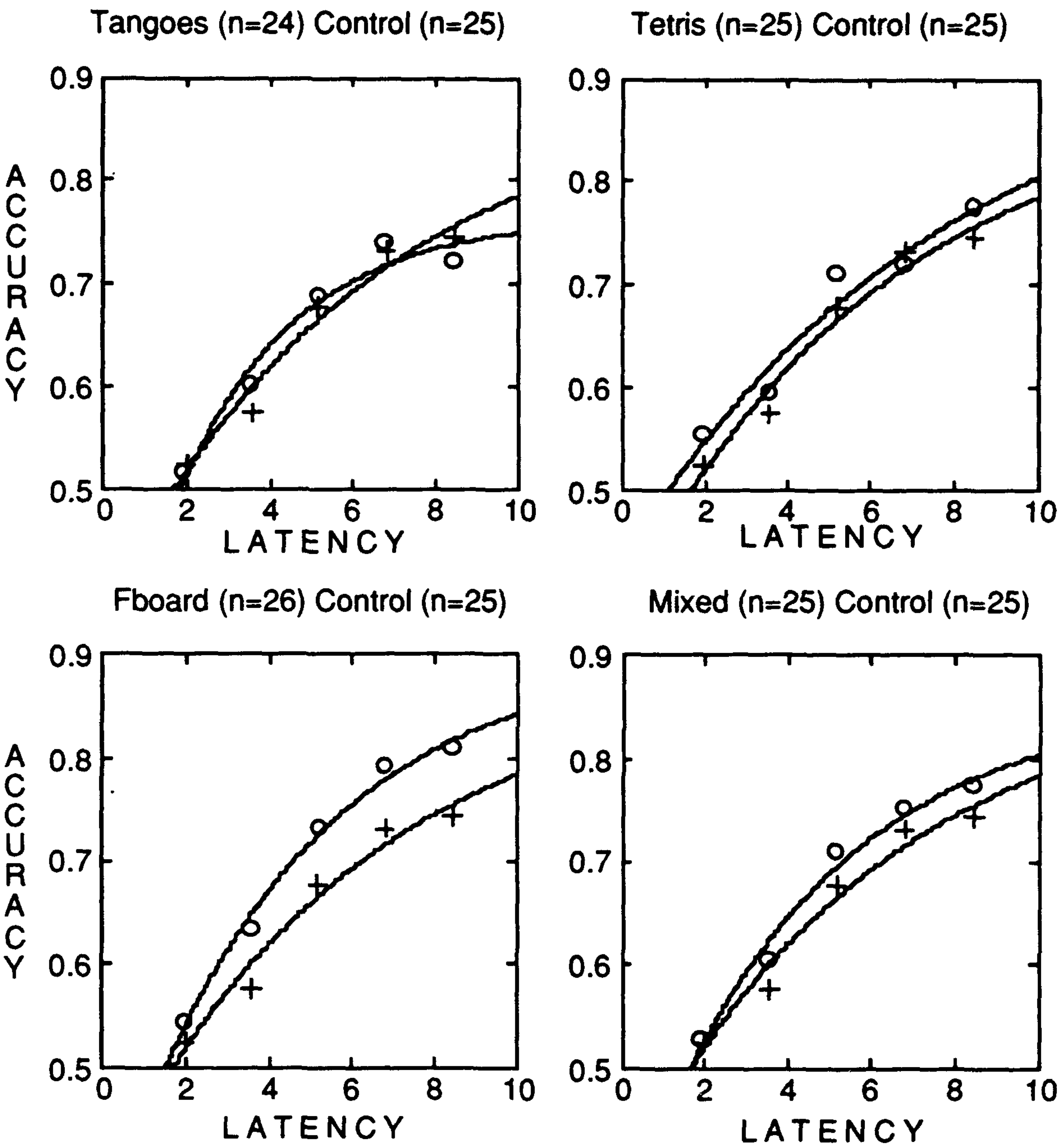


Figure 14. Accuracy-latency curves for the four practice conditions (circles), each compared to the no-practice control (crosses).

The first question we sought to address in this experiment was whether subjects in the Form Board condition who practiced on varied sets of paper-and-pencil assembly problems would show the same level of procedural transfer observed in Experiment 1 when subjects practiced an assembly task that differed from posttest only in whether were items were self-paced or presented for fixed exposures. Figure 14 shows that the Form Board group clearly outperformed other treatment groups. Thus, transfer was not entirely contingent on identity of task procedures and task formats. Nevertheless, it was slightly reduced from the level previously obtained. Average effect size across the five levels of stimulus exposure was .61 SD for the Form Board condition. The corresponding effect size for the assembly posttest in Experiment 1 was .82 SD (see Figure 11). Although based on somewhat different populations, variability of scores was approximately the same in both groups (perhaps due to careful screening of the data), and so the reduction in effect size is probably a reasonable estimate of the total contribution of specific transfer due to identical task procedures and a common, computer-based format for practice and testing. From the standpoint of ability training, these results are encouraging. From the standpoint of the test administrator, how ver, they are somewhat discouraging. Subjects need not practice the test to show large gains on it. Practice on a similar test -- even in a different format -- will do.

The second question concerned whether practice would produce merely task-specific improvements or if a more general improvement in the ability to assemble forms might be observed. On the one hand, McClelland's (1979) model claims that this sort of general improvement would result in a steeper curvature. Theories of skill development that depend on stimulus generalization (e.g., Anderson, 1983) predict that subjects in the Mixed group would be most likely to show general improvement. Figure 14 shows that these subjects did outperform the controls. However, as in the Form Board condition, the curvature parameter was not affected. Indeed, the simplest explanation is that the slight improvement reflects one session of practice on the form board task. On the other hand, internalization theories of skill acquisition (e.g., Vygotsky, 1962; Piaget & Inhelder, 1967) suggest that subjects who receive the most hands-on practice with concrete materials (Tangoes) would be most likely to show general improvements in mental assembly skill. Table 6 shows that the Tangoes group had a much steeper curvature parameter than the other groups. Therefore, practice with concrete materials seemed to be the only condition that produced a general improvement in rate of assembly. However, examination of Figure 14 suggests that this dramatically different curvature parameter is based on less than dramatic data. Thus, the result needs replication. Other attempts to teach spatial problem solving using concrete materials (e.g., Olson & Bialystok, 1983) suggest that some subjects profit from the training whereas others are impeded by it. Our small sample size permits no such comparisons, although future research might profitably attend to them.

We also predicted that subjects who practiced the Tetris game would be most likely to acquire skills required to respond rapidly on the computer-based posttest. Table 6 shows that, as McClelland's model predicts, the intercept was smallest for this group. Thus, subjects who practiced an unrelated computer game probably acquired specific skills that enabled them to respond more quickly than subjects who did not receive this practice.

Finally, we predicted that subjects in the Form Board and Mixed groups who had practiced assembling forms in stimulus Set X would show a higher asymptotic levels of accuracy on posttest problems using stimuli in Set X than posttest problems using stimuli in Set Y. Table 6 shows that the two stimulus sets were not exactly parallel (as evidenced by the small main effects for stimulus set on the intercept and asymptote parameters). However, neither the subjects in Form Board condition (who practiced assembling these forms for three sessions) nor subjects in the Mixed Condition (who practiced assembling them for one session) showed better performance on these stimuli in the posttest. This replicates the finding in Experiment 1 that subjects who practiced assembling the same stimuli showed no advantage over those who had practiced assembling different stimuli (see Figure 10). Those who practiced rotating these stimuli did show an effect, both on the assembly posttest (Figure 10) and on the rotation posttest (Figure 12).

Discussion

The results of this second experiment were generally in accordance with predictions. Nevertheless, there were several surprises and interesting leads for future research. First, it is of some interest that practice assembling concrete objects resulted in a more general improvement in assembly skill than did varied practice on different types of assembly tasks. This touches on a fundamental issue in the nature of the assembly process. Extended practice with concrete materials should be more effective if spatial assembly is acquired through internalization of an external activity. Models of cognition that assume spatial procedural knowledge is fundamentally the same as other forms of procedural knowledge may thus be at least partially incorrect. Alternatively, it could be that the tasks in the Mixed condition were too dissimilar. Subjects would then have developed task-specific procedures for each rather than general procedures that applied to all. In either case, this issue needs more study than it has received.

The second interesting finding is that practice on a paper-and-pencil assembly task on which items used varied procedures transferred to a computer-administered posttest using somewhat different procedures. Although large, effects of practice were somewhat smaller than obtained in Experiment 1 when practice task was more nearly identical to posttest task. As in that experiment, practiced subjects showed more efficient and more effective processing. More efficient processing was evidenced by higher levels of accuracy on brief exposure trials. More effective processing was evidence by a lower error rate on long exposure trials. These sorts of transfer effects show that training is not entirely task-specific. They also suggest that examinees need not have access to the test to obtain an advantage on it.

The third interesting finding here is that subjects who practiced the computer game Tetris for three sessions showed only a slight reduction in intercept. It could have been otherwise. Some have suggested that examinees who have practiced on video or computer games have an advantage over those who not spent much time with these amusements, especially on computer-based tests that require rapid responding. Although acceptance of the null hypothesis does not constitute evidence, one might have expected more transfer for subjects in the Tetris group. The practice task bore at least a family resemblance to the test task, both were administered on the same computers, and both required subjects to respond rapidly to events that they could only partially control. The key here was probably that subjects gave different responses in the two tasks. Learning a complex key-stroke procedure for manipulating stimuli would surely transfer to another task that required similar keystrokes (Singley & Anderson, 1989). Thus, the way to reduce advantage on speeded tests due to video game practice might be to keep responses different and simple.

The last interesting result was a replication of the first experiment that showed practice in assembling figures seemed to confer no advantage, whereas practice in rotating them did. This experiment showed that the failure to obtain an effect was not due to the brevity of practice. Three sessions of practice in assembling forms in this experiment gave no better results than did one session of practice in the previous experiment. Once again, the null hypothesis cannot be construed as evidence. Practice may have been more beneficial had stimuli been less familiar, or if subjects had practiced remembering and producing the stimuli rather than trying to combine simpler forms to create them. Indeed, as noted in the discussion of the first experiment, such practice may have encouraged subjects to maintain a representation of the overall shape as an assemblage of several simpler shapes rather than as a unified figure.

General Conclusions

These studies and the earlier Ackerman-Lohman (1990) experiment showed that practice has diverse but predictable effects on performance. First, practice resulted in substantial improvements in performing the same transformation on the same stimulus set. Effect sizes were approximately .8 SD for long exposure trials in both the assembly and rotation tasks. Although substantial, these effects were somewhat smaller than those obtained in studies that did not control for speed-accuracy tradeoff. Second, these studies showed that improvements also transferred to tasks that shared the same procedures. It was argued that practice clarified task demands, released working memory resources for image construction and retention, and thereby improved the overall efficiency and effectiveness of information processing. Further, the second experiment showed that procedures need not be identical in practice and test tasks in order to obtain substantial transfer. Subjects who practiced a varied set of paper-and-pencil assembly problems showed an average advantage of .6 SD over control subjects on a computer-based posttest. Furthermore, there was some evidence that general improvements in spatial assembly skill may be better produced through practice with concrete materials than through varied practiced on different assembly tasks. Finally, both experiments showed that subjects who practiced rotating stimuli later assembled or rotated them with greater efficiency than nonpracticed stimuli. However, practice in assembling stimuli produced no such effects. Improvements in processing efficiency (or rate) due to stimulus familiarity were small compared to the much larger improvements in processing effectiveness (or accuracy) due to procedural familiarity. In summary, these studies suggest extreme caution in interpreting scores on spatial tests for subjects who have had differential experience with the procedures they employ. The studies also suggest that practice on spatial tasks is not entirely task-specific; indeed, with a careful match of subject and treatment, general improvements in spatial skills may even be possible.

References

Ackerman, P. L., & Lohman, D. F. (1990). An investigation of the effects of practice on the validity of spatial tests (Final Report, Contract N66001-88-G 0291). Minneapolis, MN: Personnel Decisions Research Institute.

Alderton, D. L. (1988). Development and evaluation of Integrating Details: A complex spatial problem solving test. Unpublished working paper. Navy Personnel Research and Development Center.

Alderton, D. L., Pellegrino, J. W., & Lydiatt, S. (1984). Effects of extended practice on spatial processing ability. Unpublished manuscript, University of California, Santa Barbara.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Balke-Aurell, G. (1982). Changes in ability as related to educational and occupational experience. Goteborg Studies in Educational Sciences, 40.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity, familiarity, and individual differences. Journal of Experimental Psychology: Human Perception and Performance, 14, 12-23.

Brown, F. R. (1954). "The effect of an experimental course in geometry on ability to visualize in three dimensions" (Doctoral Dissertation, University of Illinois). Dissertation Abstracts, 15, 83.

Churchill, B. D., Curtis, J. M., Coombes, C. H., & Hassell, T. M. (1942). The effect of engineer school training on the surface development test. Educational and Psychological Measurement, 2, 279-280.

Cooper, L. A. (1982). Strategies for visual comparison and representation: Individual differences. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 1, pp. 77-124). Hillsdale, NJ: Erlbaum.

Dosher, B. A. (1984). Degree of learning and retrieval speed: Study time and multiple exposures. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 541-575.

Embretson, S. E. (1987). Improving the measurement of spatial aptitude by dynamic testing. Intelligence, 11, 333-358.

Faubian, R. W., Cleveland, E. A., & Hassell, T. W. (1942). "The influence of training on mechanical aptitude test scores." Educational Psychological Measurement, 2, 91-94.

Krumboltz, J. D., & Christal, R. E. (1960). Short-term practice effects in tests of spatial aptitude. Personnel and Guidance Journal, 384-391.

Kyllonen, P. C., Lohman, D. F., & Snow, R. E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. Journal of Educational Psychology, 76, 130-145.

Lajoie, S. (1987). Individual differences in spatial ability: A computerized tutor for orthographic projection tasks. (Doctoral Dissertation, Stanford University, 1986). Dissertation Abstracts International, 47, 3370A.

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 4, pp. 181-248). Hillsdale, NJ: Erlbaum.

Lohman, D. F. (1989). Estimating individual differences in information processing using speed-accuracy models. In R. Kanfer, P. Ackerman, & R. Cudeck (Eds.), The Minnesota symposium on learning and individual differences: Abilities, motivation, and methodology (pp. 119-163). Hillsdale, NJ: Erlbaum.

Lohman, D. F., & Nichols, P. D. (1990). Training spatial abilities: Effects of practice on rotation and synthesis tasks. Learning and Individual Differences, 2, 67-93.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. Psychological Review, 86, 287-330.

McNichol, D. (1972). A primer of signal detection theory. London: George Allen & Unwin.

Mittelholtz, D. J. (1985). The interactive effect of practice and item complexity on spatial synthesis. Unpublished masters thesis, The University of Iowa, Iowa City, IA.

Olsen, D. R., & Bialystok, E. (1983). Spatial cognition. Hillsdale, NJ: Erlbaum.

Paterson, D. G., Elliot, R. M., Anderson, L. D., Toops, H. A., & Heidbreder, E. (1930). Minnesota mechanical ability tests. Minneapolis: University of Minnesota Press.

Pellegrino, J. C. (1985, April). Information processing and intellectual ability. Paper presented at the meeting of the American Educational Research Association, New Orleans.

Piaget, J., & Inhelder, B. (1967). The child's conception of space. New York: W. W. Norton.

Regian, J. W., & Pellegrino, J. W. (1984). Practice and transfer effects in two-dimensional mental rotation. Unpublished manuscript, University of California, Santa Barbara, CA.

Salomon, G. (1974). Internalization of filmic operations in relation to individual differences. Journal of Educational Psychology, 66, 499-511.

Sanden, J. M. M. van der, & Schouten, A. (1985). Learner strategies and aptitude treatment interactions on psychomotor tasks in technical schools. Paper presented at AERA, Chicago.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Singley, M. K., & Anderson, J. R. (1989). The transfer of cognitive skill. Cambridge, MA: Harvard University Press.

Stern, W. (1914). The psychological method of testing intelligence (G. M. Whipple, Trans.). Baltimore: Warwick & York. (Original work published 1912)

Thomas, E. A. C. (1974). The selectivity of preparation. Psychological Review, 81, 442-464.

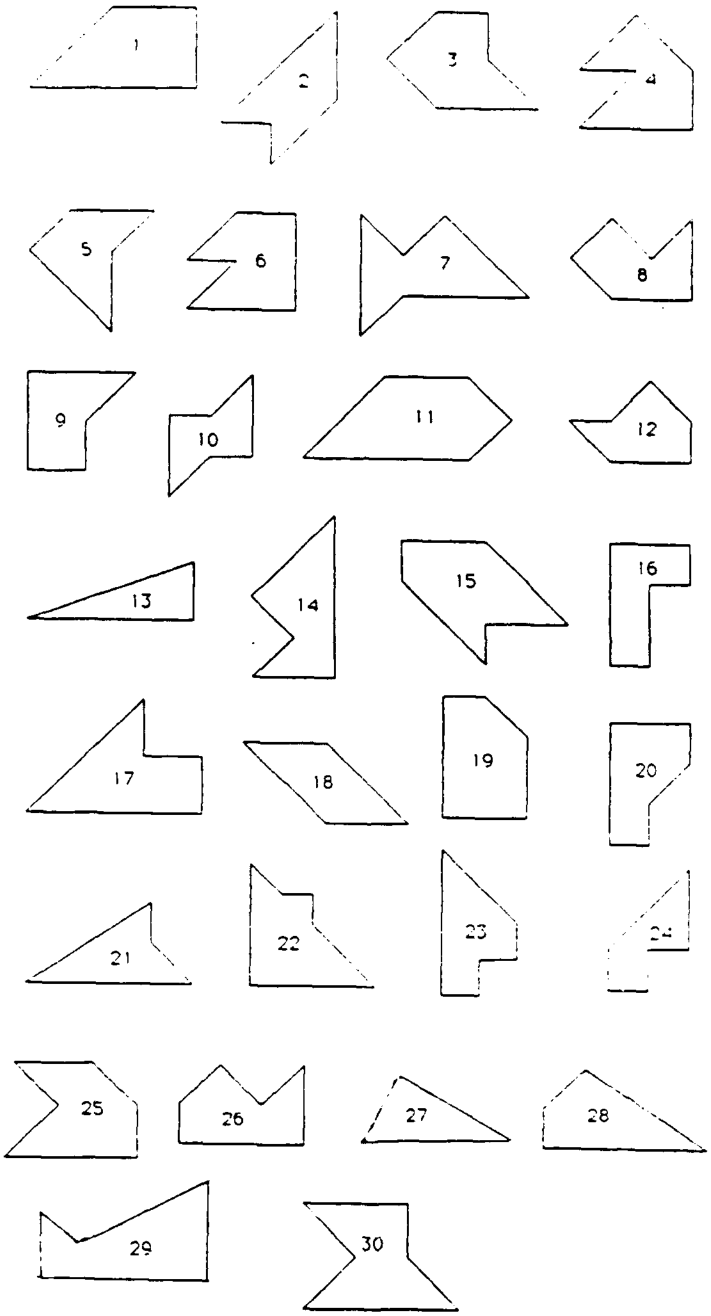
Thorndike, E. L., Bregman, E. O., Cobb, M. V., & Woodyard, E. (1926). The measurement of intelligence. New York: Columbia University, Teachers College.

Vygotsky, L. (1962). Thought and language. Cambridge, MA: MIT Press.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica, 41, 67-85.

Appendix A

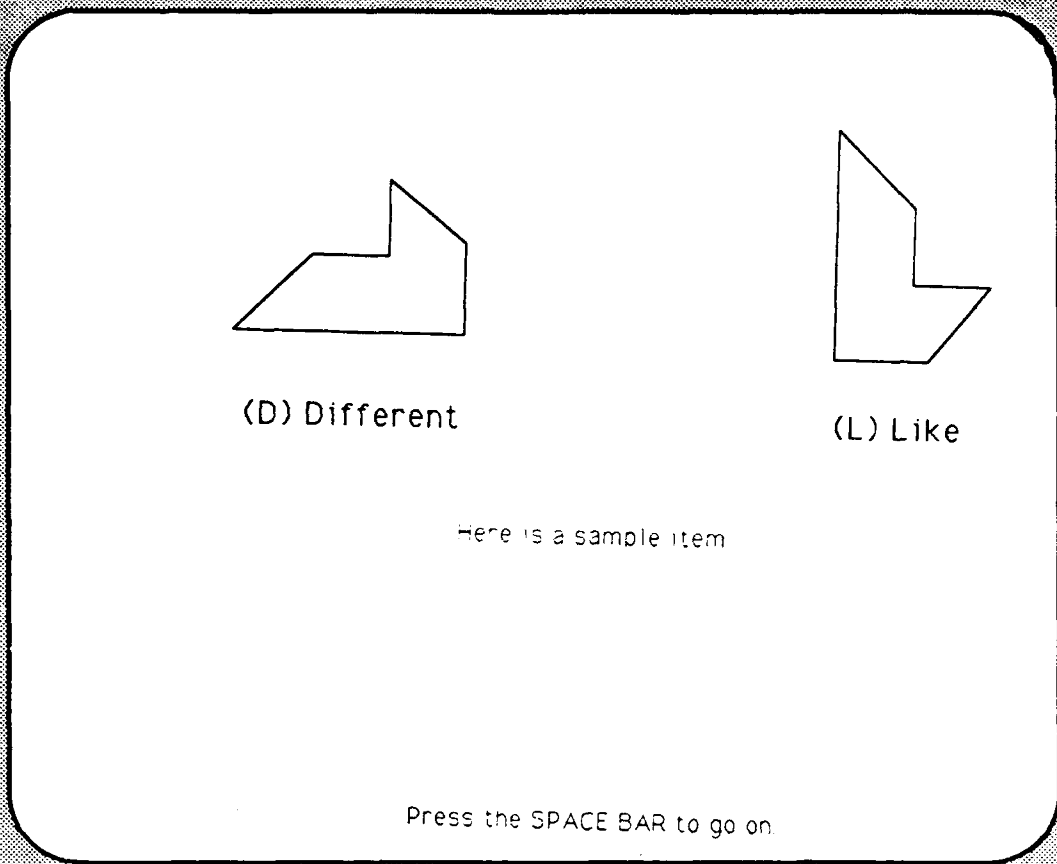
Polygon Stimuli



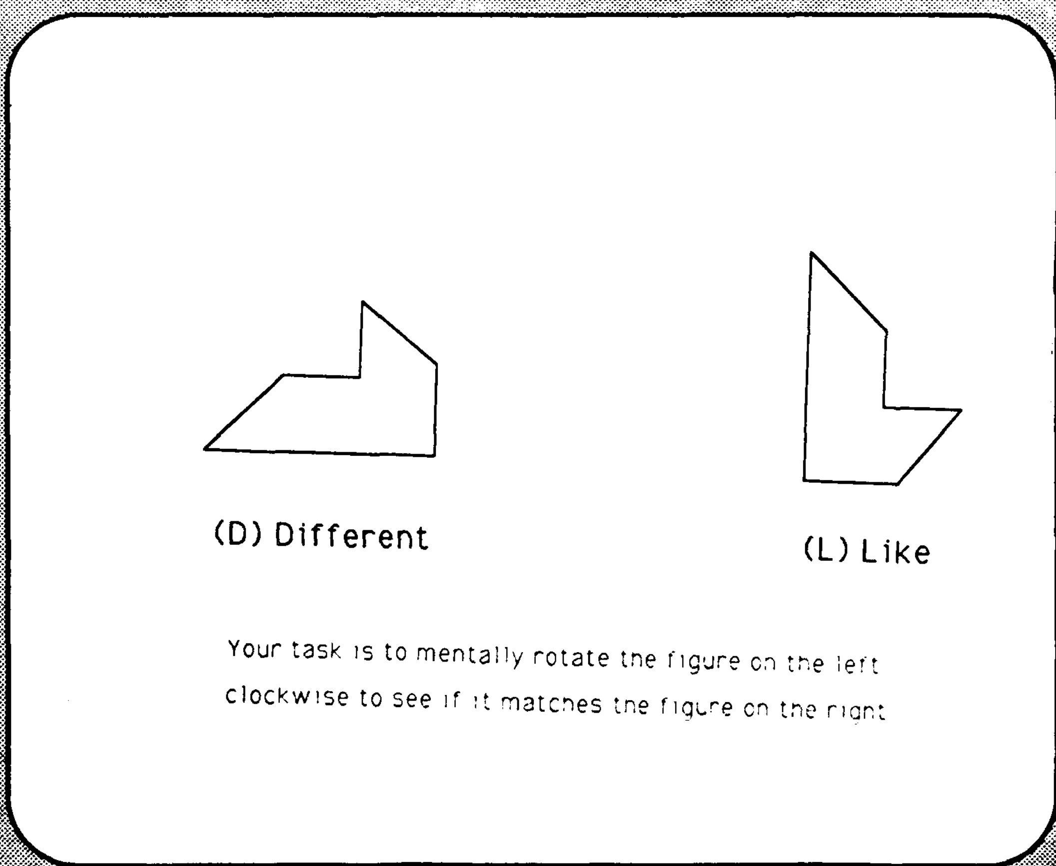
Set A;: 2, 3, 6, 7, 9, 10, 14, 16, 17, 20, 21, 24, ,25, 26, & 28 (complement in B)
Set X: 1, 3, 5, 6, 9, 10, 11, 13, 15, 16, 24, 26, 28, 29, & 30 (complement in Y)

Appendix B.1

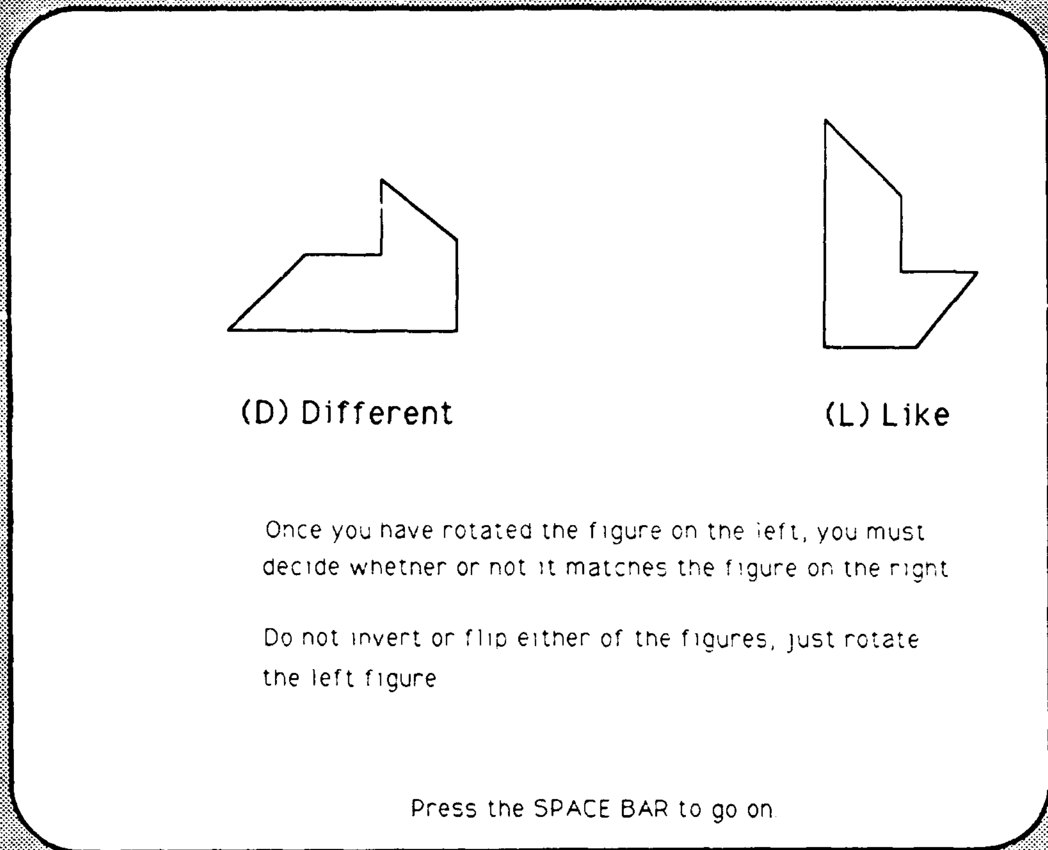
1



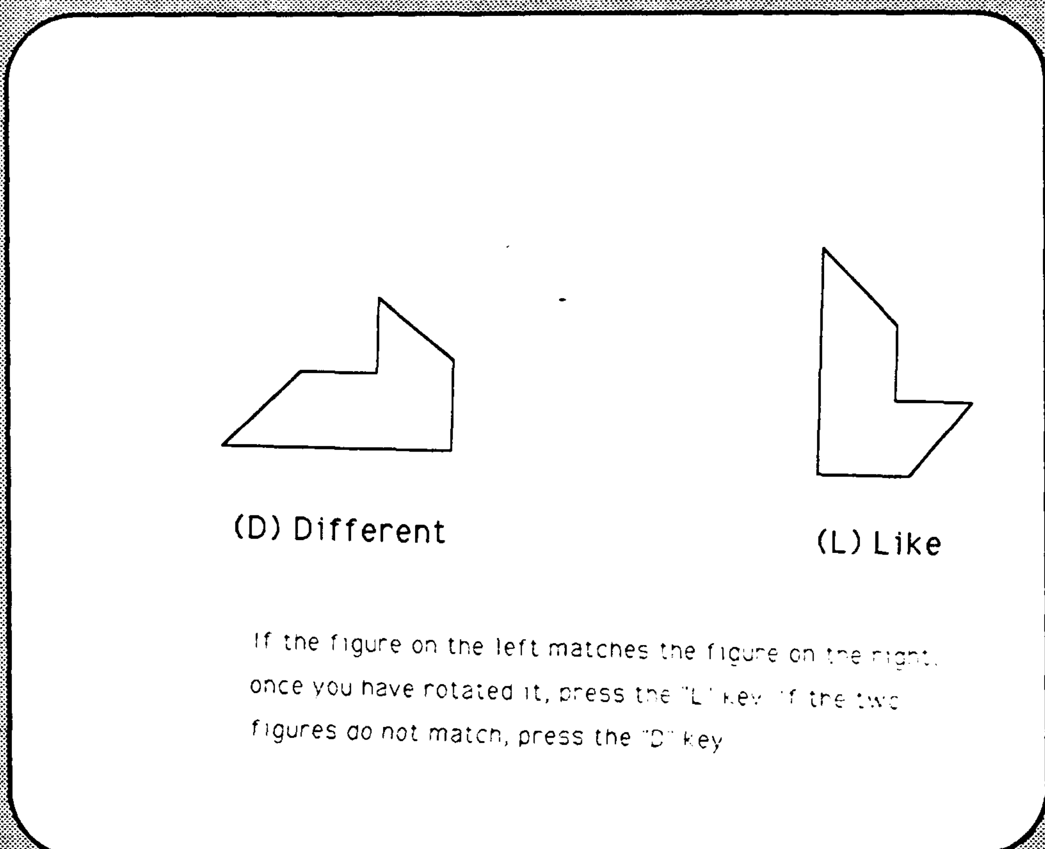
2



3


(D) Different
(L) Like
Once you have rotated the figure on the left, you must decide whether or not it matches the figure on the right
Do not invert or flip either of the figures, just rotate the left figure
Press the SPACE BAR to go on

4


(D) Different
(L) Like
If the figure on the left matches the figure on the right, once you have rotated it, press the "L" key. If the two figures do not match, press the "D" key

5

(D) Different

(L) Like

For this item, you would press the "L" key When you rotate the figure on the left, it matches the figure on the right

Press the SPACE BAR to go on.

6

(D) Different

(L) Like

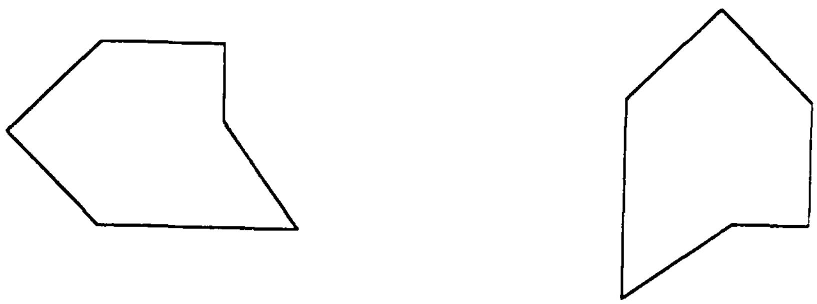
For this item you would press the "D" key When you rotate the figure on the left, it does not match the figure on the right

7

You will be able to practice on these types of items for the next several minutes. You may take as much time as you need to respond to each item, but do not take more time than is necessary.

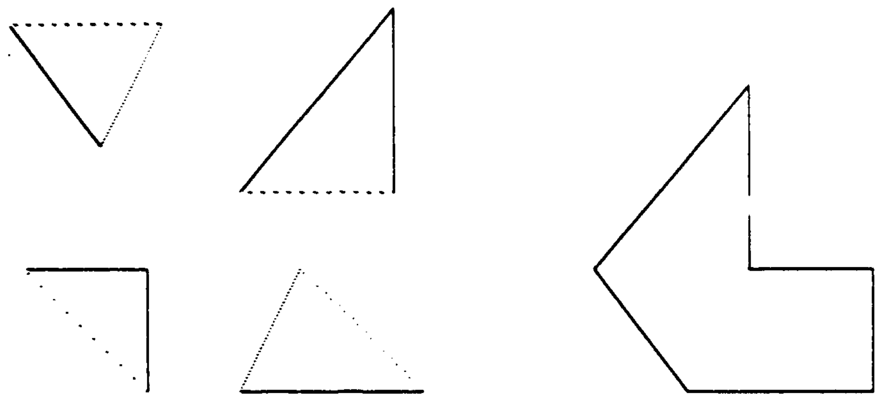
Press the SPACE BAR to go on

8



Press the SPACE BAR to go on

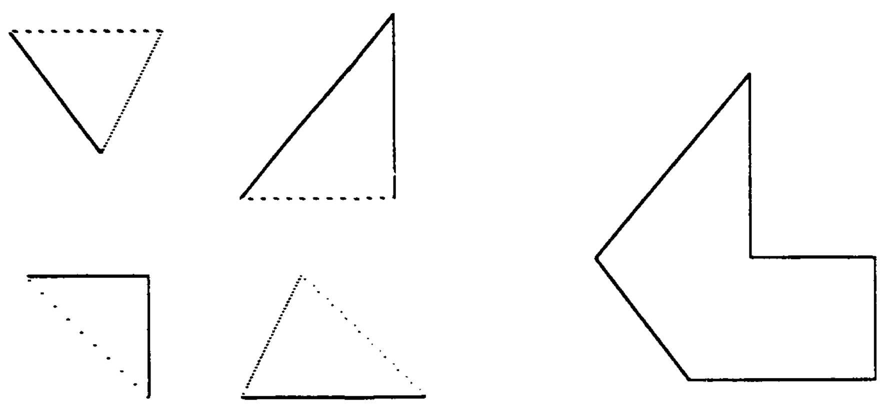
1



Here is a sample item.

Press the SPACE BAR to go on.

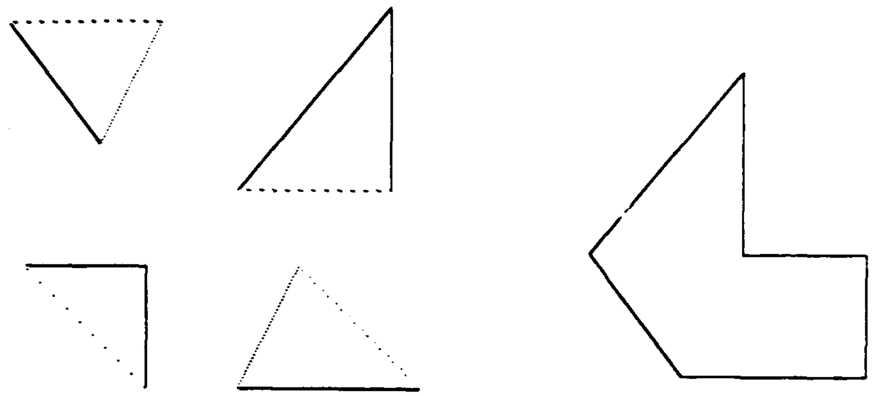
2



Notice how the pieces on the left are color coded. The colors tell you where to join the pieces together. Join the sides with same color.

Press the SPACE BAR to go on.

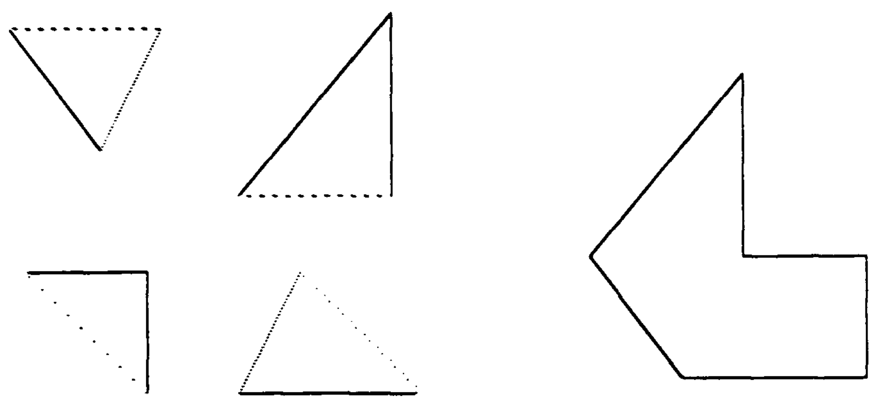
3



You will first see a figure such as this one. Your task is to mentally put together the pieces on the left, and decide if they make the figure on the right.

Press the SPACE BAR to go on.

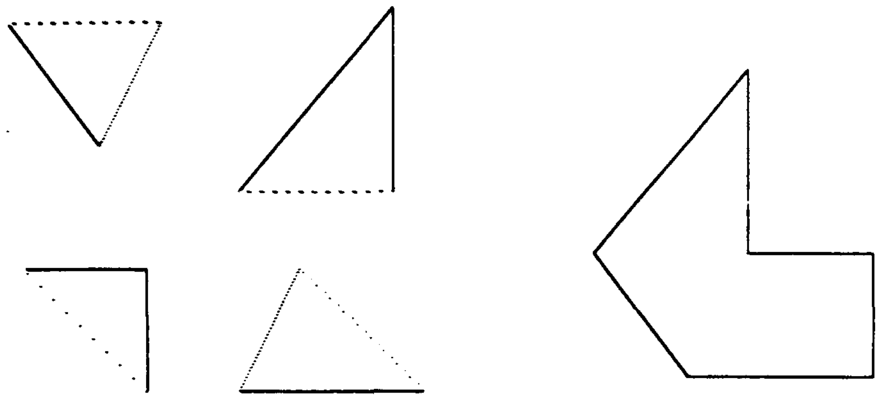
4



In this example, you would put together the two green sides, the two blue sides, and the two red sides.

Press the SPACE BAR to go on.

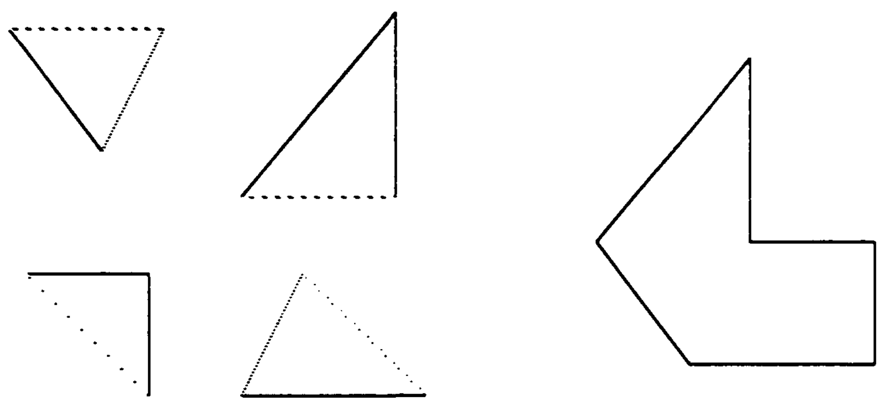
5



Do not match the gray sides, and do not flip or rotate either the target figure on the right or the pieces on the left.

Press the SPACE BAR to go on.

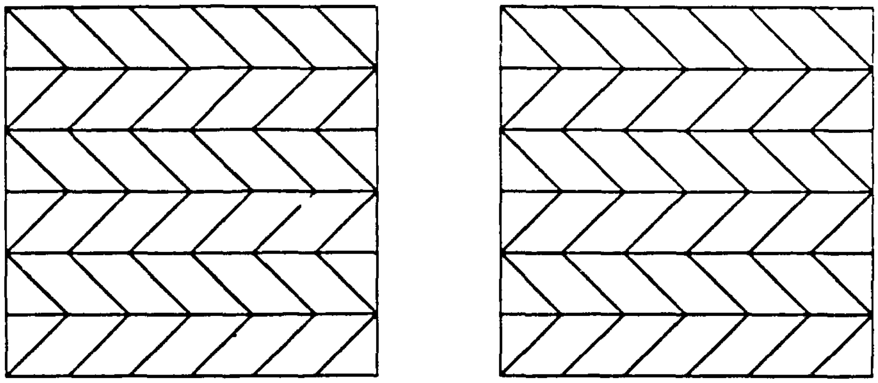
6



Do not use your finger to trace the outline of the figure on the screen. You should assemble the figures in your head, not on the screen.

Press the SPACE BAR to go on.

7



Following the figure, you will see a "mask" like this. The purpose of this mask is to prevent you from thinking about the previous figure when it no longer on the screen.

Press the SPACE BAR to go on.

8

Respond now

(D) Different **(L) Like**

After you see the mask, you will see the message above.

Once you see the message above, you must respond immediately.

If the pieces on the left DO make the figure on the right, type "L" for "Like". If the pieces on the left DO NOT make the figure on the right, type "D" for "Different".

Press the SPACE BAR to go on.

9

Respond now

(D) Different **(L) Like**

For the figure you just saw, you would press the "L" key, because the pieces on the left DID make the figure on the right. The "L" and "D" kyes are not active yet, this is just an example.

Press the SPACE BAR to go on.

10

Respond now

(D) Different **(L) Like**

You must respond very quickly for the items following these instructions. You will have only three-quaters of a second to respond once you see the message above. Do not respond now, this is just an example. Keep the middle fingers of each hand on the "D" and "L" keys (without pressing the keys down) so that you can respond as quickly as possible.

Press the SPACE BAR to go on.

Appendix C

Sample Form-Board Practice Materials

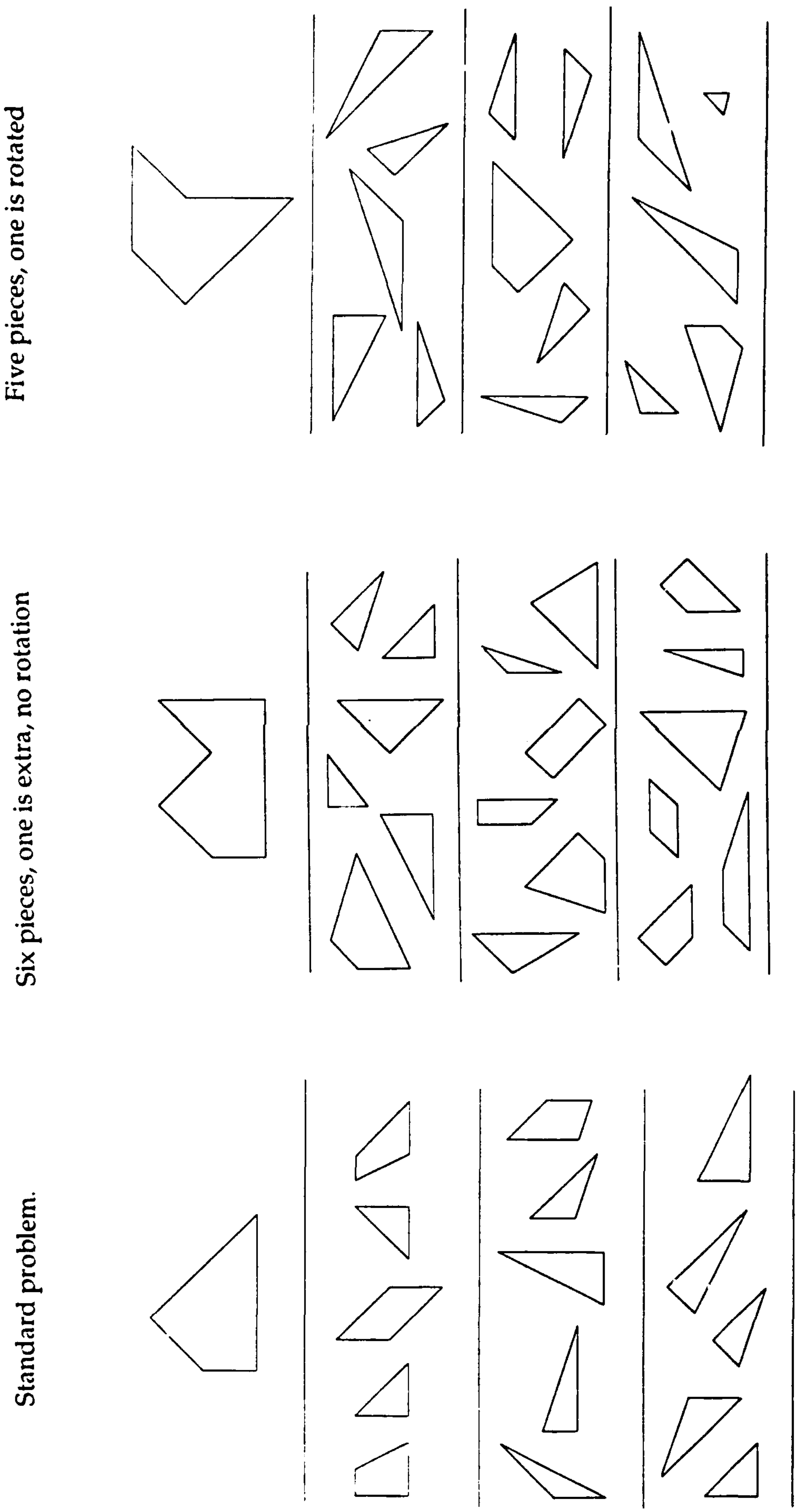
Three 15-page practice booklets were constructed. A target figure was shown at the top of each page. Each of the 15 stimuli in Stimulus Set A appeared as the target figure once in each booklet. Three rows of smaller figures were shown below the target figure. Each row contained 5 or 6 figures, depending on the task. The task was to show how the smaller figures could be combined to form the target figure. For the first eight pages of each booklet, the task was to number the sides of the five component figures to show how they should be joined to create the target figure. For the next seven pages of each booklet, the task was more complex. In Booklet 1, six component shapes were shown for each problem, only five of which were needed to construct the target figure. In Booklets 2 and 3, five component pieces were again presented. However, one had been rotated to a new position. In all problems, the subject's task was to number the sides of the component figures to indicate how they should be combined. Subjects checked their answers after completing a page of problems. They were also encouraged to try to solve problems in their heads, and not to draw lines or figures on the booklets unless absolutely stuck. A proctor provided assistance and feedback as needed.

Sample Problems

C.1 Standard problem. Five pieces, all must be used, no rotation

C.2 Six pieces, one is extra, no rotation

C.3 Five pieces, one is rotated


Five pieces, one is rotated
Six pieces, one is extra, no rotation
Standard problem.

Appendix D

Example Tangoe Item